Forecasting Stock Returns with Large Dimensional Factor Models

A. Giovannelli — University of L'Aquila, Italy

D. Massacci — King's College London

S. Soccorsi

Department of Economics

Lancaster University Management School

June 23, 2021

Abstract

We study equity premium out-of-sample predictability by extracting the information contained in a high number of macroeconomic predictors via large dimensional factor models. We compare the well-known factor model with a static representation of the common components with the Generalized Dynamic Factor Model, which accounts for time series dependence in the common components. Using statistical and economic evaluation criteria, we empirically show that the Generalized Dynamic Factor Model helps predicting the equity premium. Exploiting the link between business cycle and return predictability, we find accurate predictions also by combining rolling and recursive forecasts in real-time.

JEL classification: C38, C53, C55, G11, G17.

Keywords: Stock Returns Forecasting, Factor Model, Large Data Sets, Forecast Evaluation.

[0]This paper benefits from comments from participants at the CORE@50 Conference (Louvain-la-Neuve, Belgium), the 36th International Symposium on Forecasting (Santander, Spain) and the Macroeconomics and Financial Time Series Analysis workshop (Lancaster, UK) and from conversations with Marc Hallin and Marco Lippi. Errors and omissions are the authors' responsibility.

E-mail addresses: alessandro.giovannelli@univaq.it (Alessandro Giovannelli); daniele.massacci@kcl.ac.uk (Daniele Massacci, corresponding author); s.soccorsi@lancaster.ac.uk (Stefano Soccorsi).

1 Introduction

How should stock return prediction models incorporate the information contained in a large number of predictors that exhibit a factor structure? To the very best of our knowledge, this issue has not been addressed in the literature. We answer this question by comparing three classes of large dimensional factor models: in estimating the underlying factor space, these models differ from each other in how they account for the time series dependence in the common components. This allows us to study to what extent the information contained in the time series dimension of the common components helps predicting stock returns. This paper thus contributes to the literature on predictability of financial returns with a large number of predictors: see Elliott and Timmermann (2016).

Forecasting stock returns plays a key role in several areas of finance such as asset pricing, portfolio allocation and evaluation of investment managers performance: see Rapach and Zhou (2013) for a review of the literature. However, this is a challenging task: as discussed in Timmermann (2008), equity premium predictability is short-lived due to traders' searches for forecasting patterns. Early contributions conclude that out-of-sample predictability is either confined to specific periods (Pesaran and Timmermann, 1995) or completely absent (Bossaerts and Hillion, 1999; Goyal and Welch, 2003; Welch and Goyal, 2008). More recent evidence shows that returns are predictable by macroeconomic and financial variables (Campbell and Thompson, 2008; Rapach *et al.*, 2010; Ferreira and Santa-Clara, 2011; Pettenuzzo *et al.*, 2014; Pettenuzzo and Ravazzolo, 2016; Pan *et al.*, 2020), and by technical indicators (Neely *et al.*, 2014).

The majority of existing contributions study equity premium out-of-sample forecasting using a small set of predictors (see Rapach and Zhou, 2013): for example, the Welch and Goyal (2008) dataset is made of 14 and 15 variables at monthly and quarterly frequency, respectively. However, there is clear evidence of comovement and latent factor structure in large datasets of stock returns: these returns can be decomposed into common and idiosyncratic components, which are mutually orthogonal at all leads and lags; common components are driven by a small number of latent common factors, which determine comovements in the data.[1] This paper studies equity premium out-of-sample forecasting using a high number of macroeconomic predictors to estimate the factors driving the comovements in returns.

Early work on factor models considered small-scale datasets: Geweke (1977), and Sargent and Sims

[1]See Connor and Korajczyk (1986, 1988).

(1977), employ *exact factor models*, which impose the restriction of no cross-sectional dependence on the idiosyncratic terms. On the other hand, large dimensional factor models, pioneered by Chamberlain and Rothschild (1983), rely upon an *approximate factor structure*, in which the idiosyncratic terms are allowed to exhibit some degree of cross-sectional dependence. More recent contributions thus study large scale information sets: see Connor and Korajczyk (1986, 1988), Forni *et al.* (2000), Bai and Ng (2002), Stock and Watson (2002*a*, *b*), Forni *et al.* (2005) and Forni *et al.* (2015, 2017).

We focus on the following three classes of large-dimensional factor models, which differ from each other in how they account for time series dependence in the common components, in the estimation strategy, and in the forecasting equation; within each class, we then consider several specifications, which include different numbers of common factors.

(*a*) Stock and Watson (2002*a*) estimate common factors by principal components and compute predictions as projections onto the factor space. Based only on contemporaneous covariances, this is a static method for factor estimation and predictions are computed using a *static representation*, in which the factors are loaded contemporaneously.

(*b*) Forni *et al.* (2005) also compute predictions in a static way as projections onto the factor space. However, they allow for a data generating process with a *dynamic representation* known as the Generalized Dynamic Factor Model (henceforth GDFM), in which the common factors are loaded dynamically via one-sided filters (Forni *et al.*, 2000).

(*c*) Forni *et al.* (2015, 2017) extend the dynamic method of Forni *et al.* (2005) by allowing for an infinite-dimensional factor space: this relaxes any restriction on the lead-lag relationships among the variables and common factors, and allows for a *dynamic forecasting equation.* In this sense, Forni *et al.* (2015, 2017) provide a fully fledged dynamic approach to the estimation of the GDFM.

Existing evidence on stock returns predictability with large factor models is limited to Stock and Watson (2002*a*) static method. Ludvingson and Ng (2007) find evidence of predictability in quarterly returns using a large number of macroeconomic and financial variables. At monthly frequency, Neely *et al* (2014) conclude that Welch and Goyal (2008) low-dimensional dataset provides valuable information to predict returns when it is augmented with technical indicators. Baetje and Menkoff (2016) find that predictability stemming from Welch and Goyal (2008) dataset is unstable and declining over time.

Çakmaklı and van Dijk (2016) successfully exploit large macroeconomic information to predict monthly returns via factor augmented regressions; in a similar exercise, Gonçalves *et al.* (2017) find statistically significant predictability for some of the estimated factors. Ohno and Ando (2018) propose factor augmented regressions based on a shrinkage estimator.

None of the above mentioned contributions assesses the performance of the GDFM in predicting stock returns. Forni *et al.* (2018) extract factors from a large macroeconomic dataset similar to the one we consider in this work: their results show that the GDFM often yields more accurate predictions of macroeconomic variables than the commonly used factor model based on the static approach. Motivated by this encouraging result, we fill a gap in the stock return forecasting literature by contributing with the very first evidence of predictability based on the GDFM. Forni *et al.* (2018) focus on forecasting macroeconomic variables: our work crucially differs from theirs in that equity premium predictability patterns tend to be short-lived due to traders behavior and thus difficult to identify, as previously discussed.

We use the monthly FRED-MD large dimensional macroeconomic database of McCracken and Ng (2016) to conduct a pseudo real-time one-step-ahead equity premium forecasting exercise. We consider several forecasting methods (Giacomini and White, 2006; Timmermann, 2008) comprising aspects such as: the specification of the factor model (Stock and Watson, 2002a; Forni *et al.*, 2005; Forni *et al.*, 2015, 2017); recursive or rolling estimation windows (Timmermann, 2008); statistical and economic evaluation criteria (Leicht and Tanner, 1991; Pesaran and Timmermann, 1995). In order to facilitate comparison with the existing literature and assess the role of the macroeconomic information contained in our large dataset, we also consider the updated small-dimensional Welch and Goyal (2008) monthly dataset.

We obtain three main results. First, the information contained in large macroeconomic datasets leads to more accurate predictions both in statistical and economic terms: factor models estimated using the large-dimensional McCracken and Ng (2016) database outperform those that employ the small-dimensional Welch and Goyal (2008) dataset, as well as a range of small and medium-sized datasets obtained via a LASSO-driven variable selection. Second, predictions based on the GDFM, either by the estimator of Forni *et al.* (2005), or by that of Forni *et al.* (2015, 2017), prevail over those based on the static method of Stock and Watson (2002*a*). Third, we propose a novel method selection criterion that selects the best performing method in pseudo real-time and exploits the well known cyclicality in stock returns predictability (Rapach *et al.*, 2010)[2]: this allows us to pick a model within a given class at each point in time and to timely switch between estimation windows depending on the phase of the business cycle.[3] We check the robustness of our findings when real-time data, as opposed to revised FRED-MD data, are used to estimate the factors. Our results are qualitatively unaffected by data revisions.

Finally, we study the linkages between statistical and economic measures of forecast accuracy (Leicht and Tanner, 1991; Pesaran and Timmermann, 1995). We consider a risk-averse investor with mean-variance preferences and relative risk aversion parameter γ (see Rapach and Zhou, 2013, and references therein). Our results favour the factor models of Forni *et al.* (2005), and Forni *et al.* (2015, 2017); they also show that statistical and economic measures of forecast accuracy are generally positively correlated (Cenesizoglu and Timmermann, 2012), and that the strength of the correlation increases with γ.

The remainder of the paper is organized as follows. Section 2 explains how we forecast the equity premium with the latent factor models we consider. Section 3 describes the data. For each model, Section 4 shows the making of estimated factors, that is the contribution of each variable and how this changes over time. Section 5 assesses the out-of-sample predictive ability of the factor models. Section 6 provides two sets of additional findings: it reports real-time results and compares them with their pseudo real-time counterparts; it presents inferential results on the temporal pattern of forecast accuracy. Finally, Section 7 concludes.

2 Forecasting with latent factor models

Due to the curse of dimensionality, high-dimensional modelling is a challenge for standard parametric frameworks. Latent factor modelling turns dimensionality from a curse into a blessing: it exploits the idea that the bulk of the dynamics in the data concentrates into a few latent factors, which can be recovered by aggregating an increasing number of variables of interest. The factor models we consider differ in the way such aggregation is done.

Let $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{nt})'$ be a panel of covariance stationary time series x_{it} (with cross-section $i = 1, \ldots, n$, and time $t = 1, \ldots, T$), $\boldsymbol{\Gamma}_k = E\boldsymbol{x}_t\boldsymbol{x}'_{t-k}$ its covariance matrix with lag time k, and $\boldsymbol{\Sigma}(\theta)$

[2] Pesaran and Timmermann (2005) discuss the role of automated selection in forecasting. Timmermann (2008) stresses the importance of monitoring local predictability patterns for successful out-of-sample forecasting of stock returns.

[3] See Pesaran and Timmermann (1995), and Bossaerts and Hillion (1999), for early contributions on model selection as applied to stock returns forecasting. Clark and McCracken (2009) provide analytical, Monte Carlo, and empirical evidence of the benefits of combining estimation windows in the presence of structural breaks.

its spectral density matrix at frequency $\theta \in [-\pi, \pi]$. Define $\{v_j, z_j\}_{j=1}^n$ and $\{\lambda_j(\theta), p_j(\theta)\}_{j=1}^n$ the eigenvalues (sorted in decreasing order) and the corresponding eigenvectors of $\mathbf{\Gamma}_0$ and $\mathbf{\Sigma}(\theta)$, respectively. Factor models imply the orthogonal decomposition

$$x_{it} = \chi_{it} + \xi_{it},$$

where χ_{it} is x_{it}'s *common* component in the sense that is driven by common factors, and ξ_{it} is its *idiosyncratic* component. Since the dynamics of the common components are driven by relatively few latent factors, the number of parameters in factor models does not increase with n. Consistent estimation is typically achieved as $n \to \infty$.

As the two components are mutually orthogonal at all leads and lags, the same decomposition holds true for both $\mathbf{\Gamma}_k$ and $\mathbf{\Sigma}(\theta)$, that is

$$\begin{aligned} \mathbf{\Gamma}_k &= \mathbf{\Gamma}_k^{\chi} + \mathbf{\Gamma}_k^{\xi}, \\ \mathbf{\Sigma}(\theta) &= \mathbf{\Sigma}^{\chi}(\theta) + \mathbf{\Sigma}^{\boldsymbol{\xi}}(\theta), \end{aligned}$$

where $\mathbf{\Gamma}_k^{\chi}$ and $\mathbf{\Gamma}_k^{\xi}$ are common and idiosyncratic covariances, and $\mathbf{\Sigma}^{\chi}(\theta)$ and $\mathbf{\Sigma}^{\xi}(\theta)$ are common and idiosyncratic spectral densities.

Approximate factor structures are inferred both in the time and in the frequency domain. In fact, as $n \to \infty$ we have:

(i) the number $r \ll n$ of *static* common factors corresponds to the number of diverging eigenvalues v_j of $\mathbf{\Gamma}_0$ (Bai and Ng, 2002);

(ii) the number $q \ll n$ of *dynamic* common factors is equal to the number of spectral eigenvalues $\lambda_j(\theta)$ diverging almost everywhere in $[-\pi, \pi]$ (Hallin and Liska, 2007).

In the same way, as $n \to \infty$, idiosyncrasy is characterized by bounded idiosyncratic eigenvalues and spectral eigenvalues.[4]

In the rest of the paper, we refer to a *dynamic estimation method* for models with dynamic factors

[4]Therefore, limited amount of cross-sectional dependence between the idiosyncratic terms is allowed. This is the distinctive feature between the approximate factor models described here and the *exact* factor model studied by Geweke (1977), and Sargent and Sims (1977). In addition, serial correlation in the idiosyncratic terms is not dismissed.

estimated considering the common-idiosyncratic decomposition of the spectral density matrix and therefore account for the whole covariance structure of the data: to this category belong the factor models described in subsections 2.2 and 2.3 below. On the other hand, the *static estimation method* refers to models employing static factors, which are estimated considering contemporaneous covariances only (i.e., $\mathbf{\Gamma}_0$), rather than the whole covariance structure of the data: to this category belongs the factor model described in subsection 2.1 below.

The most general factor model, namely the GDFM, involves the following *dynamic representation* for the common components

$$\chi_{it} = \frac{\boldsymbol{c}_i(L)}{\boldsymbol{d}_i(L)}\boldsymbol{f}_t, \tag{1}$$

where $\boldsymbol{c}_i(L)$ and $\boldsymbol{d}_i(L)$ are one-sided polynomials in the lag operator L with square-summable coefficients, and $\boldsymbol{f}_t$ is a $q-$dimensional orthonormal white noise (see Forni *et al.* 2000) . The main advantage with respect to competing factor models is that, beyond stationarity and regularity conditions for the existence of spectral density matrix $\mathbf{\Sigma}(\theta)$, the GDFM does not require assumptions on the dynamics of the factor structure to achieve consistent estimation of the common components. On the contrary, the widespread *static representation*

$$\chi_{it} = \boldsymbol{\lambda}_i'\boldsymbol{F}_t, \tag{2}$$

where $\boldsymbol{\lambda}_i$ are factor loadings and the factors $\boldsymbol{F}_t$ are possibly serially correlated, imposes strong restrictions on the data generating process if consistency is to be achieved (see Hallin and Lippi, 2013; Forni *et al.* 2015, 2017). Nevertheless, factor models with static representation should still be considered dynamic time series models because they can accommodate some forms of dynamics. For instance,

$$\chi_{it} = \quad a_{i1}\underbrace{\frac{f_{1t}}{1-\alpha L}}_{F_{1t}} \quad + \quad a_{i2}\underbrace{f_{2t}}_{F_{2t}} \quad + \quad a_{i3}\underbrace{f_{2t-1}}_{F_{3t}} \tag{3}$$

allows for a static representation with three static factors (F_{1t}, F_{2t}, F_{3t}) and a dynamic representation with two dynamic factors (f_{1t}, f_{2t}). On the contrary,

$$\chi_{it} = \quad a_i\frac{f_{1t}}{1-\alpha_i L} = \quad a_i\left(f_{1t} + \alpha_i f_{1t-1} + \alpha_i^2 f_{1t-2} + \dots\right) \tag{4}$$

does not allow for a (finite-order) static representation. The model in subsection 2.3 below is based on the dynamic representation (1); those in subsections 2.1 and 2.2 below are based on the static representation (2).

In the following three subsections we outline the predictive methods considered and their corresponding forecasting equations. In so doing, without any loss of generality, we assume that the excess return — i.e. our target, which we label ρ_t (see Section 3) — is the first variable in our vector of observables. Formally, $x_{1t} = \rho_t$.

2.1 Static method, static representation (SW)

The method proposed by Stock and Watson (2002*a*), henceforth SW, involves static principal components and projections on the factor space. Let $\widehat{\mathbf{\Gamma}}_0$ be the sample counterpart of $\mathbf{\Gamma}_0$. Static factors are extracted from $\boldsymbol{x}_t$ by taking the r principal components of $\widehat{\mathbf{\Gamma}}_0$ that solve the eigenvalue problem

$$\widehat{z}_j \widehat{\mathbf{\Gamma}}_0 = \widehat{v}_j \widehat{z}_j, \qquad j = 1, \ldots, r.$$

The estimated factors are $\boldsymbol{F}_t^{SW} = \widehat{\boldsymbol{z}} \boldsymbol{x}_t$, where $\widehat{\boldsymbol{z}} = (\widehat{z}_1, \ldots, \widehat{z}_r)'$, and the $h-$step ahead forecast of ρ_t is

$$\widehat{\rho}_{t+h|t}^{SW} = \widehat{\boldsymbol{\beta}} \boldsymbol{F}_t^{SW}, \tag{5}$$

where $\widehat{\boldsymbol{\beta}}$ is a r-dimensional row vector of projection coefficients onto the space spanned by $\boldsymbol{F}_t^{SW}$.

2.2 Dynamic method, static representation (FHLR)

The dynamic method proposed by Forni *et al.* (2005), henceforth FHLR, is a two-step procedure based on the dynamic estimation method and predictions formed from a static representation via a *constrained* projection onto the factor space.

Step one: estimation

The spectral density matrix of the data at frequency $\theta \in [-\pi, \pi]$ is estimated through discrete Fourier transforms of the sample covariance matrix

$$\widehat{\mathbf{\Sigma}}(\theta) = \frac{1}{2\pi} \sum_{k=-M}^{M} e^{-ik\theta} w_k \widehat{\mathbf{\Gamma}}_k,$$

where w_k are the weights of a window function and M is a truncation parameter.[5]

Letting $\widehat{p}_j(\theta)$ and $\widehat{\lambda}_j(\theta)$ be the eigenvector and eigenvalues of $\widehat{\boldsymbol{\Sigma}}(\theta)$, the spectral density matrices of the common and idiosyncratic components are computed as

$$\widehat{\boldsymbol{\Sigma}}^{\boldsymbol{\chi}}(\theta) = \sum_{j=1}^{q} \widehat{\lambda}_j(\theta)\,\widehat{p}_j(\theta)'\,\widehat{p}_j(\theta), \qquad (6)$$

$$\widehat{\boldsymbol{\Sigma}}^{\boldsymbol{\xi}}(\theta) = \sum_{j=q+1}^{n} \widehat{\lambda}_j(\theta)\,\widehat{p}_j(\theta)'\,\widehat{p}_j(\theta), \qquad (7)$$

respectively; the covariances via inverse Fourier transforms are

$$\widehat{\boldsymbol{\Gamma}}_k^{\chi} = \frac{2\pi}{2H+1}\sum_{j=-H}^{H} e^{ik\theta}\widehat{\boldsymbol{\Sigma}}^{\boldsymbol{\chi}}(\theta_j), \qquad (8)$$

$$\widehat{\boldsymbol{\Gamma}}_k^{\xi} = \frac{2\pi}{2H+1}\sum_{j=-H}^{H} e^{ik\theta}\widehat{\boldsymbol{\Sigma}}^{\boldsymbol{\xi}}(\theta_j), \qquad (9)$$

with Fourier frequencies $\theta_j = \frac{2\pi j}{2H+1}$.

Step two: forecasting equation

The so-called *generalized principal components* of the couple $\left(\widehat{\boldsymbol{\Gamma}}_0^{\chi}, \widehat{\boldsymbol{\Gamma}}_0^{\xi}\right)$ solve the eigenvalue problem

$$\widehat{z}_j^g \widehat{\boldsymbol{\Gamma}}_0^{\chi} = \widehat{v}_j^g \widehat{z}_j^g \widehat{\boldsymbol{\Gamma}}_0^{\xi}, \qquad j = 1, \ldots, r,$$

$$\text{subject to} \quad \begin{cases} \widehat{z}_j^g \widehat{\boldsymbol{\Gamma}}_0^{\xi} \widehat{z}_j^g = 1 \\ \widehat{z}_i^g \widehat{\boldsymbol{\Gamma}}_0^{\xi} \widehat{z}_j^g = 0, \quad i \neq j. \end{cases}$$

Letting $\widehat{\boldsymbol{z}}^g = (\widehat{z}_1^g, \ldots, \widehat{z}_r^g)'$ be a vector of the first r generalized eigenvectors, $\boldsymbol{F}_t^{FHLR} = \widehat{\boldsymbol{z}}^g \boldsymbol{x}_t$ is the vector of estimated factors.[6] Notice that this requires $r < \infty$, so just like the static one of SW, this estimator becomes inconsistent if the data generating process does not admit a finite dimensional static factor representation – like e.g. the model (4).

[5] All empirical results in Section 5 are obtained using a standard triangular window.

[6] Generalized eigenvectors correspond to eigenvectors of data weighted according to their signal to noise ratio (Forni *et al.*, 2005).

Finally, the $h-$step ahead forecast of ρ_t is given by

$$\widehat{\rho}_{t+h|t}^{FHLR} = \widehat{\boldsymbol{\delta}} \boldsymbol{F}_t^{FHLR}, \tag{10}$$

where $\widehat{\boldsymbol{\delta}}$, equal to the first row of the $n \times r$ matrix $\widehat{\boldsymbol{\Gamma}}_h^{\chi} \widehat{\boldsymbol{z}}^{g\ \prime} \left(\widehat{\boldsymbol{z}}^g \widehat{\boldsymbol{\Gamma}}_0 \widehat{\boldsymbol{z}}^{g\prime} \right)^{-1}$, is an r-dimensional row vector of *constrained* projection coefficients onto the space spanned by $\boldsymbol{F}_t^{FHLR}$. Notice that such *constrained* projection imposes dynamic factor structure restrictions through $\widehat{\Gamma}_h^{\chi}$ rather than $\widehat{\Gamma}_h$ as in the unconstrained projection (5) employed by SW.

2.3 A fully fledged dynamic method (FHLZ)

The method proposed by Forni *et al.* (2015, 2017), henceforth FHLZ, shares with FHLR the decomposition of the spectral density matrix in (6) and (7), and that of the covariances in (8) and (9), estimated as in Step one described in subsection 2.2.

Letting $\boldsymbol{\chi}_t^{(i)}$ be any $q+1-$dimensional subvector of common components, according to (1) it has a common factor moving average representation $\boldsymbol{\chi}_t^{(i)} = \frac{\boldsymbol{c}^{(i)}(L)}{\boldsymbol{d}^{(i)}(L)} \boldsymbol{f}_t$, where $\frac{\boldsymbol{c}^{(i)}(L)}{\boldsymbol{d}^{(i)}(L)}$ is a $(q+1) \times q$-dimensional filter. Forni *et al.* (2015, 2017) prove that, since moving average representations with such "tall" filters — i.e. with more rows than columns — are *generically fundamental*[7], they can be inverted into an autoregressive representation

$$\boldsymbol{A}^{(i)}(L) \boldsymbol{\chi}_t^{(i)} = \mathbf{R}^{(i)} \boldsymbol{f}_t,$$

where $\boldsymbol{A}^{(i)}(L)$ is $(q+1) \times (q+1)$, $\mathbf{R}^{(i)}$ is $(q+1) \times q$, and the lag order of $\boldsymbol{A}^{(i)}(L)$ is finite and can be suitably determined.[8] Let us stack all $q+1-$dimensional vectors of common components: we thus obtain an autoregressive representation in which the dynamic factors $\boldsymbol{f}_t$ are loaded only contemporaneously in $\boldsymbol{A}^{(i)}(L) \boldsymbol{\chi}_t^{(i)}$.[9] The dynamic factors can then be consistently estimated via principal components of filtered data

$$\boldsymbol{S}_t = \underline{\boldsymbol{A}}(L) \boldsymbol{x}_t = \underline{\boldsymbol{R}} \boldsymbol{f}_t + \underline{\boldsymbol{A}}(L) \boldsymbol{\xi}_t,$$

[7] More precisely, they are invertible because tall filters are generically zeroless. On the other hand, non-zeroless moving averages admit a multitude of nonfundamental representations which cannot be inverted into causal vector autoregressive representations (e.g., see Soccorsi, 2016). The genericity argument means that such property holds everywhere in the parameter space apart from a measure zero subset.

[8] In Section 4 results are obtained by determining the lag order of $\boldsymbol{A}^{(i)}(L)$ via a BIC information criterion.

[9] In order to avoid heavier notation, we are assuming without loss of generality that $g = n/(q+1)$; as discussed by Forni *et al.* (2015, 2017) no special challenge arises when n is not a multiple of $q+1$.

where $\boldsymbol{S}_t$ collects the stacked vectors $\boldsymbol{A}^{(i)}(L)\boldsymbol{\chi}_t^{(i)}$, $\underline{\boldsymbol{R}}$ is a tall $n \times q$ matrix and the $n \times n$ autoregressive filter takes the form

$$\underline{\mathbf{A}}(L) = \begin{pmatrix} \boldsymbol{A}^{(1)}(L) & 0 & \dots & 0 \\ 0 & \boldsymbol{A}^{(2)}(L) & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \dots & \boldsymbol{A}^{(g)}(L) \end{pmatrix}.$$

Denote by $\widehat{\boldsymbol{S}}_t$ the estimator for $\boldsymbol{S}_t$. Let $(\omega_1, \dots, \omega_q)$ be the first q eigenvalues of the covariance matrix of $\widehat{\boldsymbol{S}}_t$, and $\Psi = (\psi_1, \dots, \psi_q)'$ be a $q \times n$ matrix collecting the associated eigenvectors: the estimated dynamic factors are $\boldsymbol{f}_t^{FHLZ} = \Psi \widehat{\boldsymbol{S}}_t$. The estimated autoregressive coefficients of $\underline{\mathbf{A}}(L)$ are computed from the estimated common covariances $\widehat{\boldsymbol{\Gamma}}_k^{\chi}$ in equation (8) and $\underline{\widehat{\boldsymbol{R}}} = \Psi$. Given these quantities, we estimate impulse responses to the dynamic factors as

$$\widehat{\mathbf{w}}(L) = \underline{\widehat{\boldsymbol{A}}}(L)^{-1}\,\underline{\widehat{\boldsymbol{R}}} \quad ,$$

where its generic element $\widehat{\mathbf{w}}_i$ is consistent for $\frac{\mathbf{c}_i(L)}{\mathbf{d}_i(L)}$ in (1), for any $i \in [1, n]$. Finally, recalling that ρ_t is the first element of the n-dimensional vector $\boldsymbol{x}_t$, $h-$step ahead predictions of excess returns are:

$$\widehat{\rho}_{t+h|t}^{FHLZ} = \widehat{\mathbf{w}}_{1h}\boldsymbol{f}_t^{FHLZ} + \widehat{\mathbf{w}}_{1h+1}\boldsymbol{f}_{t-1}^{FHLZ} + \dots \quad , \tag{11}$$

where $\widehat{\mathbf{w}}_{1h}$ is a q-dimensional row vector of projection coefficients onto the space spanned by $\boldsymbol{f}_t^{FHLZ}$.

3 Data

Letting T_e be the end point of the estimation period, our aim is to produce one-step-ahead out-of-sample forecasts for the sequence of stock returns at each point in time $\tau+1$ given the information available at time τ, for $\tau = T_e, \dots, T-1$. We use monthly data on stock returns along with a large set of predictors from which we estimate the factors: these are 122 variables included in the FRED-MD database described in McCracken and Ng (2016). Our data sample spans the period January 1960 to December 2019. We also use the 14 predictors originally proposed in Welch and Goyal (2008), and subsequently extended up to 2019 by the same authors: this allows for comparison with existing studies using low dimensional sets of predictors, which are reviewed in Rapach and Zhou (2013).

Stock returns are computed from the S&P 500 index in excess of a short T-bill rate and include dividends. Formally, we let ρ_{t+1} be the excess return at period $t+1$, for $t = 1, \cdots, \tau - 1$: the goal is to produce one-step-ahead out-of-sample forecasts of $\rho_{\tau+1}$ given the information set available at time τ, for $\tau = T_e, \ldots, T-1$.

The FRED-MD database organizes the variables into eight groups: (*i*) output and income; (*ii*) labor market; (*iii*) consumption and orders; (*iv*) orders and inventories; (*v*) money and credit; (*vi*) interest rate and exchange rates; (*vii*) prices; (*viii*) stock market. The choice of the 122 variables was based on data availability over the period of interest as reported in the Appendix.

The 14 predictors proposed in Welch and Goyal (2008) are: log dividend-price ratio (log (DP)), log dividend-yield (log (DY)), log earnings-price ratio (log (EP)), log dividend-payout ratio (log (DE)), stock variance (SVAR), book-to-market ratio (BM), net equity expansion (NTIS), treasury bill rate (TBL), long-term yield (LTY), long-term return (LTR), term spread (TMS), default yield spread (DFY), default return spread (DFR), lagged inflation (INFL).[10] As discussed in Pettenuzzo *et al.* (2014), the predictors fall into the following broad categories: (*i*) valuation ratios (log (DP), log (DY), log (EP), BM); (*ii*) measures of bond yields (TBL, LTY, TMS, DFY, DFR); (*iii*) estimates of equity risk (LTR, SVAR); corporate finance variables (log (DE), NTIS); (*iv*) macroeconomic variables (INFL).

Table 1 about here

Table 1 provides summary statistics for the series of excess stock returns and for the variables included in the Welch and Goyal (2008) dataset. Despite the difference in the sample period of interest, the figures are aligned to those displayed in Table 1 in Pettenuzzo *et al.* (2014).[11]

As argued in Hansen and Timmermann (2012), a crucial issue in out-of-sample forecasting exercises is the choice of the sample-split between estimation and evaluation periods to avoid data mining. Following Timmermann (2008), we use the first 10 years of data as a training sample and we evaluate the forecasts over the period January 1970 to December 2019: a long evaluation sample allows for stronger power of forecast evaluation tests and minimizes the likelihood of spurious rejections. The end point T_e of the

[10] As in Welch and Goyal (2008), we lag inflation by an extra period to allow for the delay in CPI releases.
[11] Table 1 in Pettenuzzo *et al.* (2014) covers the sample period January 1927 to December 2010.

estimation window thus is December 1969.

4 A preliminary look at the factors

We now look at the role that each variable in the FRED-MD dataset, as described in Section 3, plays in estimating the factors through the three models discussed in Section 2. We do so through the Degree of Commonality (hereafter DC): this measures the share of the variance of x_{it} explained by the common factors. Formally, DC is defined as

$$DC_{it}^{(m)} = \frac{var\left(\chi_{it}^{(m)}\right)}{var\left(x_{it}\right)}, \qquad i \in [1, n], \ t \in [T_e + 1, T], \tag{12}$$

where $m \in (SW, FHLR, FHLZ)$ denotes the factor model, $var\left(\chi_{it}^{(m)}\right)$ is the variance of the common component estimated by the factor model m, and $var\left(x_{it}\right)$ is the variance of x_{it}. We calculate this measure through a rolling exercise, replicating the forecasting exercise carried out in Section 5. We present $DC_{it}^{(m)}$ for each cross-sectional unit i and time period t through the heat-maps in Figure 1: the variables are divided into the groups described in Section 3; the groups are labelled in the ordinate and separated by thick horizontal red lines.

Figure 1 about here

DC displays patterns both across groups and over time. Variables included in "output and income" and in "interest and exchange rates" contribute the most to the estimated factors in all the three factor models. "Price" variables are also relevant, especially in the second half of the sample. "Money and credit" variables become important in the second half of the sample.

Although the three models display clear similarities, some differences are worth discussing. On the one hand, the evolution of $DC_{it}^{(SW)}$ resembles that of $DC_{it}^{(FHLR)}$; on the other hand, $DC_{it}^{(FHLZ)}$ generates somehow different paths, which may be due to the unique fully fledged dynamic form of the FHLZ estimator discussed in Section 2. For example, "interest and exchange rate" variables play an important role over the whole sample period in the case of SW and FHLR factors; conversely, the same variables are important for FHLZ factors mainly in the central and final parts of the sample.

5 Out-of-sample analysis

5.1 Forecasting methodology

As in Timmermann (2008), we explicitly follow Giacomini and White (2006) and distinguish between forecasting *model* and forecasting *method*. The former refers to the underlying econometric specification, in our case the three factor models described in Section 2. The latter includes the model and other choices made by the forecaster, such as the estimator for the model unknown parameters (as discussed in Section 2), the length of the estimation window, and the evaluation criteria.

5.1.1 Econometric model

We estimate the factors from the small-dimensional Welch and Goyal (2008) dataset and from the large collection of FRED-MD variables described in Section 3. In the former case, we consider $r = 1, 2, 3$ static and $q = 1, 2$ dynamic factors. From the FRED-MD database, we estimate up to $r = 15$ and $q = 5$ static and dynamic factors, respectively: for ease of exposition and without loss of generality, we report results for $r = 1, 5, 15$, and $q = 1, 3, 5$, only.

5.1.2 Estimation window

We consider recursive window and rolling window estimation schemes. Given the sample split described in Section 3, the former uses data from 1960:01 up to the time the forecast is made to produce a series of one-step ahead forecasts: the first forecast uses data from 1960:01 to 1969:12 to obtain an out-of-sample prediction for 1970:01; the second forecast uses data from 1960:01 to 1970:01 to produce a forecast for 1970:02, and so on. As in Timmermann (2008), the rolling window scheme employs a fixed-length window of the most recent ten years of data (i.e., 120 monthly observations) to estimate the models and produce the sequence of one-step ahead forecasts. The recursive window scheme is commonly used in the empirical literature on out-of-sample stock return forecasting: see Pesaran and Timmermann (1995), Campbell and Thompson (2008), Welch and Goyal (2008), Rapach *et al.* (2010), and Pettenuzzo *et al.*, (2014). The rolling window scheme is common in the macroeconomic forecasting literature concerned with structural breaks in macroeconomic data: see Stock and Watson (2012) and Forni *et al.* (2018).

5.1.3 Evaluation criteria

As in Pesaran and Timmermann (1995), and Bossaerts and Hillion (1999), the first evaluation criterion we consider is the mean squared prediction error (MSPE), which assesses the absolute performance of a sequence of forecasts. These are produced as in equations (5), (10) and (11) for SW, FHLR and FHLZ, respectively. We next compare the forecasts obtained from the factor models in relation to a given benchmark. Following Campbell and Thompson (2008), and Welch and Goyal (2008), we take as a benchmark the prevailing mean (PM), namely

$$\rho_{t+1} = \alpha + \varepsilon_{t+1}, \quad t = 1, \ldots, \tau - 1, \quad \tau = T_e, \ldots, T-1, \tag{13}$$

where ε_{t+1} is a white noise error term with unpredictable mean. The equity premium forecast for period $\tau + 1$ made at time τ is $\hat{\rho}_{\tau+1,\text{rec}} = \tau^{-1} \sum_{t=1}^{\tau} \rho_t$ under recursive window; it is equal to $\hat{\rho}_{\tau+1,\text{rol}} = T_e^{-1} \sum_{t=\tau-T_e+1}^{\tau} \rho_t$ under rolling window. The recursive window scheme produces the benchmark usually employed in the equity premium forecasting literature: see Campbell and Thompson (2008), and Welch and Goyal (2008). As discussed in Timmermann (2008), the choice of the estimation window is a function of the underlying assumption made about the mean of the equity premium: when estimated recursively, the model in (13) assumes the equity premium has a constant mean and it is not predictable; the rolling window scheme implies that the mean of the equity premium slowly changes over time.

The MSPE may be used to measure the out-of-sample goodness of fit of a sequence of forecasts. To this purpose, we next consider the out-of-sample R^2 employed in Campbell and Thompson (2008), Timmermann (2008), Welch and Goyal (2008), Rapach *et al.* (2010), and Pettenuzzo *et al.* (2014). Let MSPE_1 and MSPE_0 be the mean squared prediction errors from any factor model and from the prevailing mean in (13), respectively: the out-of-sample R^2 is $R^2_{OoS} = 1 - \text{MSPE}_1 / \text{MSPE}_0$. By construction, $R^2_{OoS} \leq 0$ if and only if $\text{MSPE}_1 \geq \text{MSPE}_0$, meaning that the benchmark is at least as good as the alternative model at forecasting $\rho_{\tau+1}$; conversely, $R^2_{OoS} > 0$ if and only if $\text{MSPE}_1 < \text{MSPE}_0$.

Finally, we assess the statistical significance of the improvement of the alternative model over the benchmark by testing the null hypothesis $R^2_{OoS} \leq 0$ against the one-sided alternative $R^2_{OoS} > 0$. We run the Clark and West (2007) test (hereafter CW): this is robust to the different degrees of estimation error between models, which would otherwise favor the more parsimonious benchmark.

5.1.4 The role of the business cycle

Rapach *et al.* (2010), Henkel *et al.* (2011), and Rapach and Zhou (2013) argue that stock returns predictability exhibit discernible patterns linked to business cycle dynamics. We then assess our forecasts over the entire evaluation period, as well as during NBER-dated expansions and recessions.

5.2 Empirical results

5.2.1 Recursive window

Table 2 displays the out-of-sample R^2 for the recursive window scheme.

Table 2 about here

When factors are extracted from Welch and Goyal (2008) small dimensional dataset (Panel A), the best performing model is FHLZ with $q = 1$ dynamic factor: the model outperforms all other specifications over the entire evaluation period, as well as during recessions and expansions; the out-of-sample R^2 is always positive and significant at 5% level or less.[12] The out-of-sample R^2 is equal to 0.98% over the whole evaluation period, and to 0.81% and 1.39% during expansions and recessions, respectively. The forecasts produced by FHLZ with $q = 1$ dynamic factor are thus more accurate during contractionary periods: this is consistent with Campbell and Cochrane (1999), Menzly *et al.* (2004), and Bekaert *et al.* (2009), who argue that risk premia are countercyclical and drive (at least part of) predictability; it also resembles Rapach *et al.* (2010), Henkel *et al.* (2011), and Rapach and Zhou (2013), who empirically show that out-of-sample stock returns predictability increases during recessions as compared to expansions.

Panel B in Table 2 also includes results from the best performing specifications. It shows that when the whole large dimensional FRED-MD dataset is used to estimate the factors, the best performing model is FHLR. The specifications with $r = 2, 3$ static factors and $q = 1$ dynamic factor overall produce the most accurate forecasts, with statistically significant improvements over the prevailing mean at 10% level or less. These two models also outperform FHLZ with $q = 1$ dynamic factor estimated from the small-dimensional Welch and Goyal (2008) dataset (see Panel A): with the data at hand, large dimensional

[12] As stressed in the footnote 19 of Pettenuzzo *et al.* (2014), the p-values from the Clark and West (2007) test should be interpreted with caution and in line with Diebold (2015): those p-values should be intended to compare forecasts rather than models.

factor models provide a hedge over smaller scale counterparts.

Boivin and Ng (2006) question whether adding series with little factor structure to estimate factors may result in factors being less useful for out-of-sample forecasting purposes. Panels C, D and E in Table 2 show results when factors are estimated from 25, 50 and 75 series from the FRED-MD dataset, respectively: the series are selected through a pseudo real-time LASSO procedure at each time τ the forecast is made. When only a subset of the series is used to estimate the factors, the forecasting ability of the models, as measured by the out-of-sample R^2, deteriorates. The performance of the forecasts improves as more series are used to estimate the factors. This result shows the usefulness of large data and is in line with basic asymptotic results on consistent factor estimation, which is achieved for growing cross-sections.

Given the data at hand, large dimensional datasets are more informative than small-dimensional counterparts in a recursive window framework; FHLR models overall produce the most accurate forecasts with $r = 2, 3$ static factors and $q = 1$ dynamic factor.

5.2.2 Rolling window

Table A1 in Online Appendix displays the out-of-sample R^2 for the rolling window scheme.As with the recursive window, factor models estimated using the large dimensional FRED-MD dataset (see Panel B) generally produce more accurate forecasts than those based on the small scale Welch and Goyal (2008) dataset (see Panel A). Over the whole forecasting period, FHLZ forecasts are the most precise: the out-of-sample R^2 ranges between 0.88% and 1.96%, and it is always statistically greater than zero at least at 5% level. During expansionary periods, no large dimensional model unambiguously dominates any other specification: in particular, FHLZ with $q = 1$ dynamic factors as obtained from the Welch and Goyal (2008) dataset has the highest out-of-sample R^2 out of all models, which is equal to 1.09% and it is significant at 5% level. During economic recessions, FHLR and FHLZ models combined with the large dimensional FRED-MD dataset produce forecasts of similar accuracy: almost all out-of-sample R^2 are positive and significant at least at 10% level. Overall, the forecasts from SW models are less accurate. When factors are estimated from a subset of 25, 50 and 75 variables from the FRED-MD dataset selected using a LASSO type procedure (see Panels C, D and E, respectively), the quality of the forecasts deteriorates; as in the case of the recursive window scheme, the precision of the forecasts improves as more variables are used to estimate the factors.

In conclusion, the results from the rolling window scheme generally favor large dimensional factor models, with FHLR and, especially, FHLZ having a hedge over SW.

5.2.3 The role of large macroeconomic information

In order to assess the role of our large cross-section of macroeconomic information, in Table 3 we resume the LASSO results obtained for $r = 1, 2, 3, 4, 5, 10, 15$, and $q = 1, 2, 3, 4, 5$, by picking the best specifications of SW, FHLR, FHLZ for each level of variable selection (i.e., 25, 50, and 75 variables) under both estimation windows over the full sample.

Table 3 about here

Comparing the results from the cross-sections restricted by the LASSO with those from the full FRED-MD dataset, we observe an almost monotonic improvement in the number of included variables, as evidenced by the increasing out-of-sample R^2. In line with standard asymptotic results on factor models as discussed in Section 2, this empirical finding suggests that the underlying assumptions are likely to hold in the data at hand and the models we consider are not misspecified. This confirms the conclusion of McCracken and Ng (2016), who propose the FRED-MD dataset as a resource for factor analysis. Stock and Watson (2012), and Giannone *et al.* (2017), find similar results regarding the performance of shrinkage methods as applied to forecasting problems with large-dimensional macroeconomic datasets.

5.3 An adaptive method selection approach

The results discussed in Sections 5.2.1 and 5.2.2 and displayed in Tables 2 and A1 in Online Appendix, respectively, show two important findings: factor models estimated using large dimensional datasets tend to produce more reliable forecasts than those estimated using a smaller number of macroeconomic series; models based on the dynamic method (i.e., FHLR and FHLZ) outperform SW, which is based on the static method. These findings come from a high number of forecasting methods, as discussed in Section 5.1.

We also find an empirical regularity along the business cycle: rolling forecast are more accurate during recessions, while recursive forecasts have an edge during expansions. As argued in Pesaran and

Timmermann (2005), decision makers require selecting the best performing method in real-time. We thus implement what we label a *method selection criterion*: in the spirit of Pesaran and Timmermann (1995), and Bossaerts and Hillion (1999), this allows us to pick a model within a given class at each point in time; it further allows us to timely switch between estimation windows, whose importance is stressed in Clark and McCracken (2009), and Pesaran and Timmerman (2007). We can thus exploit more fully the cyclical behaviour in returns predictability discussed in Rapach and Zhou (2013).

5.3.1 Model selection strategy

In the spirit of Pesaran and Timmermann (1995), and Bossaerts and Hillion (1999), we study the model selection problem within SW, FHLR and FHLZ for a given estimation window.

When implementing model selection criteria using the recursive window scheme, we have to account for structural instability in the underlying factor model: see Baltagi *et al.* (2017), and references therein. We adopt the following strategy to tackle the problem of model selection in the presence of structural instability under the recursive window scheme. As suggested in Stock and Watson (2012), we *a priori* select $r = 5$ static factors and we keep this number fixed over the entire out-of-sample evaluation period. To the very best of our knowledge, no existing study allows us to *a priori* fix the number of dynamic factors. At each point in time, we choose $q = 4$ dynamic factors using Hallin and Liška (2007) criterion as applied to the rolling window scheme: model instability is less likely to affect this estimation scheme as the dynamic window effectively adapts to time variation in the loadings.[13]

The empirical results with $r = 5$ and $q = 4$ show that FHLR and FHLZ outperform SW: the former two produce forecasts with higher out-of-sample R^2 than the latter during the entire out-of-sample evaluation period, as well as during expansions and recessions.[14] Between the two dynamic models, FHLZ generates more accurate forecasts than FHLR over the full evaluation period and in expansionary phases: the out-of-sample R^2 is equal to 0.68% and 0.48%, respectively, and in both cases it is significant at 5% level. During recessions, the out-of-sample R^2 from FHLZ forecasts is marginally higher than that from FHLR forecasts; in the latter case, statistical significance is achieved at 10% level. Overall, FHLZ

[13]At each point in time, Hallin and Liška (2007) criterion selects between $q = 4$ and $q = 5$ dynamic factors. We choose $q = 4$ as a matter of parsimony. Results with $q = 5$ are very similar and are available upon request.

[14]The out-of-sample R^2 for FHLR with $r = 5$ and $q = 4$ is equal to -0.36, -0.98, 1.12 in the full sample, expansions and recessions, respectively. The corresponding values for FHLZ with $q = 4$ are 0.68, 0.48, 1.15, respectively. For both models, the out-of-sample R^2 computed over the full sample and in expansionary periods are significant at 5% level; for FHLR, 10% statistical significance is achieved during recesssions.

has an edge over FHLR under the recursive window scheme.

Under rolling window estimation, at each point in time we choose the number of static and dynamic factors according to Bai and Ng (2002), and Hallin and Liška (2007) criteria, respectively: the problem of structural instability is likely to be less relevant in this case, as rolling window estimation accounts for time-variation in the parameters. As with the recursive window scheme, FHLR and FHLZ fare better than SW (see Panel B in Table A1 in Online Appendix). Between the two dynamic models, FHLZ forecasts are better during the whole evaluation period and in expansionary phases: the out-of-sample R^2 is equal to 1.51% and 0.38%, respectively and, in both cases, it is significant at 5% level. Rolling forecasts from FHLR are more accurate during recessions: the higher out-of-sample R^2 is 5.24% and significant at 5% level.

5.3.2 Switching the estimation window in pseudo real-time

As shown in Paye and Timmermann (2006), and Rapach and Wohar (2006), return prediction models are subject to structural instabilities. Clark and McCracken (2009), and Pesaran and Timmerman (2007), argue that forecast accuracy in the presence of breaks may be improved by combining recursive and rolling estimation windows in such a way that optimally handles the trade-off between variance (which decreases with the sample size and so in recursive windows) and bias (which is generated by the breaks and so is less harmful within rolling windows). Inoue *et al.* (2017) develop a procedure to determine the window size in the presence of structural instability. Based on our empirical findings, we propose to select the estimation window as a function of business cycle conditions so that forecast accuracy can enhance in the presence of instabilities linked to the business cycle. This is relevant as out-of-sample stock returns predictability depends on the business cycle, as stressed in Rapach *et al.* (2010).

Table 4 about here

In order to empirically motivate our strategy, Table 4 reports mean squared prediction errors (multiplied by 100) for the three large dimensional factor models we consider under recursive and rolling window estimation (Panels A and B, respectively). We first look at the whole set of available factor models. During economic expansions, all models produce better forecasts under recursive window estimation. The scenario changes during contractionary periods: forecasts from SW models have similar

MSPE under recursive and rolling windows; FHLR and FHLZ models produce more accurate forecasts under rolling window. This finding is confirmed when the model selection strategy detailed in Section 5.3.1 is applied within each class of factor models: this *a priori* selects $r = 5$ and $q = 4$ under recursive window; it resorts to Bai and Ng (2002), and Hallin and Liška (2007) criteria under rolling window. The results confirm that FHLR and FHLZ generate better forecasts under recursive and rolling window during expansions and recessions, respectively.[15] This last finding suggests that timely switching between estimation windows may improve the quality of the forecasts.

In order to select the estimation window, we follow Banbura *et al.* (2011) and employ a nowcasting procedure that tracks the current state of the economy. At each point in time τ, we use the sequence $\{\text{ADS}_t\}_{t=1}^{\tau}$ of business cycle indicators of Aruoba *et al.* (2009) and select the estimation window by solving

$$\hat{\theta}_\tau = \arg\min_{\theta} \left| \left[\tau^{-1} \sum_{t=1}^{\tau} \mathbb{I}\left(\text{ADS}_t < \theta\right) \right] - R \right|, \quad \tau = T_e, \ldots, T-1,$$

where $\mathbb{I}(\cdot)$ denotes the indicator function and $|\cdot|$ the absolute value of the argument: $R = 0.14$ is the approximate sample frequency of recessions over the $1946:01 - 1969:12$ period as identified by the NBER business cycle dates.[16] At each point in time τ, the threshold $\hat{\theta}_\tau$ minimizes the distance between the empirical frequency R and the one identified by Aruoba *et al.* (2009) business cycle indicator; for each τ, we select recursive and rolling window if $\text{ADS}_\tau \geq \hat{\theta}_\tau$ and $\text{ADS}_\tau < \hat{\theta}_\tau$, respectively.

Table 5 about here

Results related to the proposed model selection criterion are displayed in Table 5. The table collects results for the PM model in (13), and for the large dimensional factor models SW, FHLR and FHLZ: under recursive and rolling windows, factor models are estimated as detailed in Sections 5.2.1 and 5.2.2, respectively; the model selection criterion chooses the estimation window according to the procedure previously described in this section. The results show that the mean squared prediction error (multiplied by 100) for FHLR and FHLZ is minimized when the method selection criterion is implemented and it is equal to 0.1911 and 0.1897, respectively (see Panel A): these value are lower than any other obtained

[15] See Section 5.3.1 for details.

[16] We also considered the case in which the empirical frequency of recessions at time τ is determined by the expanding window between $1946:01$ and τ. The results are quantitatively very similar to those from the case $R = 0.14$ and are available upon request.

from PM and SW. Table 6 also calculates the out-of-sample R^2 with respect to the most accurate benchmark, namely PM estimated by recursive window (Panel B): SW delivers negative values regardless of the estimation window; FHLR forecasts obtained from the method selection criterion outperform the benchmark with a positive and strongly significant out-of-sample R^2 equal to 0.65%; FHLZ produces the most accurate forecasts, which always deliver positive values for the out-of-sample R^2, with highest value equal to 1.41% achieved under the proposed method selection procedure.

5.3.3 Statistical forecast accuracy and portfolio choice

Following the pioneering work of Leicht and Tanner (1991), and Pesaran and Timmermann (1995), we finally study the economic value of equity premium forecasts. Our interest lies in understanding the linkages between statistical and economic measures of forecasting performance. This is an open issue: Leicht and Tanner (1991), and Cenesizoglu and Timmermann (2012), only find weak relationships between statistical and economic measures of forecast accuracy; at the same time, Pesaran and Granger (2000) advocate a closer link between decision theory and the forecast evaluation problem.

In line with Campbell and Thompson (2008), Rapach *et al.* (2010), Ferreira and Santa Clara (2011), Rapach and Zhou (2013) and Neely *et al.* (2014), we economically evaluate the forecasts by computing the certainty equivalent return for a risk-averse investor with mean-variance preferences and relative risk aversion parameter γ. At the end of each month, the investor allocates her wealth between stocks and a riskless asset. The choice depends on a dynamic trading strategy based on a benchmark and an alternative prediction method. As customary in the literature, our benchmark is the prevailing mean estimated with recursive window: as pointed out in Timmermann (2008), it assumes the equity premium has a constant mean and it is not predictable; it also produces the lowest mean squared prediction error out of all methods based on the prevailing mean (see Panel A in Table 5).

Formally, let $j = 0$ and $j = 1$ denote the benchmark and the alternative method, respectively. If the investor opts for method j at period $\tau + 1$, at period τ she assigns to stocks a share $w_{j\tau}$ equal to

$$w_{j\tau} = \frac{1}{\gamma} \frac{\hat{\rho}_{j,\tau+1}}{\hat{\sigma}^2_{j,\tau+1}}, \quad j = 0, 1,$$

where $\hat{\rho}_{j,\tau+1}$ and $\hat{\sigma}^2_{j,\tau+1}$ are the point forecasts of $\rho_{\tau+1}$ and of its variance $\sigma^2_{\tau+1}$ made at time τ, respectively: as in Campbell and Thompson (2008), we compute the latter as the five-year moving window of past monthly returns, so that $\hat{\sigma}^2_{j,\tau+1} = \hat{\sigma}^2_{\tau+1}$ is independent of the underlying forecasting method j. The realized return on the investment portfolio from method j at time $\tau + 1$ then is

$$R^p_{j,\tau+1} = w_{j\tau}\rho_{\tau+1} + r_{f\tau}, \quad \tau = T_e, \ldots, T-1, \quad j = 0, 1.$$

The certainty equivalent return is the average realized utility from method j over the out-of-sample period and it is defined as

$$\bar{U}_j = \bar{\mu}^p_j - \frac{1}{2}\gamma\left(\hat{\sigma}^p_j\right)^2, \quad j = 0, 1,$$

where $\bar{\mu}^p_j$ and $\left(\hat{\sigma}^p_j\right)^2$ are the sample mean and variance, respectively, of the portfolio returns $R^p_{\tau,t+1}$ over the out-of-sample period. Following Campbell and Thompson (2008), we constrain the portfolio weights $w_{0\tau}$ and $w_{1\tau}$ such that $0 \leq w_{0\tau}, w_{1\tau} \leq 1.5$. We then compute the utility gain

$$\Delta = \bar{U}_1 - \bar{U}_0 :$$

as discussed in Fleming et *al.* (2001), the utility gain represents the portfolio management performance fee that a mean-variance investor is willing to pay to switch from the dynamic trading strategy based on the benchmark to the one based on the alternative method. In the empirical application, we set $\gamma = 3, 4, 5, 10$: these are aligned to the values chosen in Rapach *et al.* (2010), Rapach and Zhou (2013), and Cenesizoglu and Timmermann (2012). We then multiply Δ by 1200 to express it in average annualized percentage returns.

The results are displayed in Panels C, D, E and F of Table 5 for $\gamma = 3, 4, 5, 10$, respectively. The recursive window estimation scheme leads to less accurate forecasts than those produced by the benchmark uniformly across models as all utility gains are negative. Interesting results arise under the rolling window scheme and the method selection criterion. The most accurate forecasts are produced by FHLR and FHLZ: the former prevails for $\gamma = 3, 4$ and $\gamma = 5$ under rolling window and method selection, respectively; the latter is preferable for $\gamma = 10$ under method selection. The empirical results indicate that the link between statistical and economic measures of forecast accuracy increases with risk aversion. An analysis of the correlation between out-of-sample R^2 and utility gains confirms this first impression: the correlation is equal to -0.41, 0.06, 0.25 and 0.61 for $\gamma = 3, 4, 5, 10$, respectively, and thus monotonically increases in γ. This empirical regularity is further illustrated in Figure 2, which plots utility gains against out-of-sample R^2 for the values of γ of interest.

Figure 2 about here

In conclusion, our results show that for the empirically relevant values $\gamma = 3, 4, 5$, FHLZ with rolling window produces the most accurate forecasts as evaluated in economic terms. In the extreme case $\gamma = 10$, the method selection criterion as applied to FHLZ is preferable. As in Cenesizoglu and Timmermann (2012), we further show that statistical and economic measures of forecast accuracy tend to be positively correlated; in addition, the strength of the correlation increases with γ.

6 Further results

6.1 Real-time evidence

The results in Section 5 are based on the FRED-MD dataset: since the macroeconomic variables in FRED-MD are observed at their final vintage after multiple revisions, those results come from a pseudo real-time forecasting exercise. As noticed in Ghysels *et al.* (2018), data revisions affect bond return predictability: we investigate whether we face a similar issue in forecasting stock returns with large dimensional factor models.

We assess the robustness of our main findings by running an analysis similar to the one in Section 5 using data that are available in real-time. We employ a dataset of 62 variables collected from the St. Louis Fed ALFRED data archive: the sample starts in February 1982 and ends in December 2019.[17] The ADS business cycle indicator we use to combine rolling and recursive forecasts according to the method selection criterion described in Section 5.3 is also subject to data revisions: for this reason, we consider its real-time version. ADS vintage data are provided by the Federal Reserve of Philadelphia and are available starting from 2009.[18] Due to real-time data availability, rolling and recursive forecasts are evaluated between February 1992 to December 2019, whereas forecast combinations obtained using the method selection criterion begin in January 2009 and finish in December 2019.

[17] ALFRED data are available at `https://alfred.stlouisfed.org`. In the construction of the dataset here employed, we discarded time series therein for which too few vintages are available together with other irregular time series. More details on the ALFRED dataset are provided in the Appendix.

[18] All ADS data are available at `https://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index`.

The sample for real-time analysis is a subsample of that available for the pseudo real-time exercise conducted in Section 5: for a meaningful comparison, we also report the results from the pseudo real-time exercise (i.e., based on the FRED-MD dataset) over the sample period that begins in 2009. Table 6 compares the results in terms of MSPE and out-of-sample R^2: panels A and B refer to the exercise with data in pseudo real-time; panels C and D show the results obtained using only information available in real-time. Recursive window forecasts are less accurate than those produced by the benchmark uniformly over all models when real-time data are involved; these results are in line with the findings from pseudo real-time data with the exception of those obtained from FHLZ, which produces enhanced forecasts with respect to the benchmark when revised data are employed. Under rolling window, SW and FHLR underperform with respect to the benchmark with real-time data, whereas FHLZ becomes more accurate. Finally, when the method selection criterion is used, real-time and pseudo real-time information produce qualitatively similar results.

Table 6 about here

The results in Table 6 let us conclude that there is no sizeable difference in predictive accuracy when real-time, as opposed to revised, data are used for factor estimation and return forecasting. It should also be noticed that ALFRED and FRED-MD are datasets of different cross-sectional dimensions, being made of 62 and 122 variables, respectively. Other conditions being equal, this difference in the cross-sectional dimension between the two datasets would *a priori* predict a relative deterioration in factor estimation when ALFRED is involved. In fact, such a deterioration is not observed in our empirical findings: unlike those in Ghysels *et al.* (2018) on bond returns, our results show that stock return predictability does not deteriorate when real-time data are employed; a possible explanation is that bond and equity markets react differently to macroeconomic news, as argued in Andersen *et al.* (2007).

6.2 Local predictability

Following Farmer *et al.* (2019), we now study local predictability, as opposed to global predictability, which was the focus of the analysis up to now: we do so using the FRED-MD data described in Section 3. In order to perform time-varying inference, we employ the test proposed by Giacomini and Rossi (2010) (hereafter GR): this allows us to investigate how the predictability provided by our forecasting methods distributes over time with respect to the benchmark. GR test results on the forecasts produced by the selected specifications of SW, FHLR and FHLZ are reported in Figure 3. The solid lines correspond to the GR test statistics, which is the normalized and smoothed difference between the square forecast errors of a given factor model, estimated with one of the methods considered, and the benchmark; the smoothing considered is the centred moving average of 60 observations corresponding to 5 years of forecast. The zero horizontal line indicates equal performance, the dotted lines indicate the 5% critical values, so that the factor model considered outperforms (underperforms) the benchmark locally in time at the 5% significance level when the solid line is below (above) the lower (upper) dashed line.

Figure 3 about here

Recursive and rolling predictions are the top and central plots, respectively. All models significantly outperform the benchmark during the first decade of our sample, regardless of whether the estimation window is recursive or rolling: this finding is consistent with Pesaran and Timmermann (1995), who find evidence of predictability during the 1970s. The remaining part of our sample is also associated with some predictability: this is obtained with the rolling window and is caught for a more sustained period by FHLZ, while it is very short-lived in the case of SW and FHLR forecasts. SW recursive and FHLR (either rolling and recursive) outperform the benchmark at some point in the late 80s, although only for a short period of time.

We present GR test results for SW, FHLR and FHLZ forecast combinations based on our method selection criterion in the bottom plots of Figure 3. By comparing our forecast combinations with recursive and rolling forecasts in the top and centre plots, respectively, we can see that the improvements within that time period are evident for all the factor models: FHLR predictions are significantly more accurate than the benchmark for a relatively long time after the crisis; our method selection criterion as applied to FHLZ produces significantly more accurate forecasts for a prolonged period despite the fact that the underlying recursive and rolling forecasts only infrequently outperform the benchmark; SW recursive and, less frequently, SW rolling are significantly outperformed by the benchmark, whereas SW combined predictions are generally as accurate as the benchmark.

7 Conclusions

We study one-step-ahead out-of-sample predictability of the monthly equity premium using large dimensional factor models. We compare the static method of Stock and Watson (2002*a,b*) with the more general approach known as Generalized Dynamic Factor Model, for which the two estimators proposed in Forni *et al.* (2005), and Forni *et al.* (2015, 2017), are considered. Through statistical and economic forecast evaluation criteria, we show that large dimensional factor models condense the information contained in a high number of predictors to accurately forecast stock returns, especially when the Generalized Dynamic Factor Model is considered. Further improvements are found by applying a combination of recursive and rolling forecasts, which we label method selection: this combines information stemming from both windows, depending on the underlying state of the economy.

Our work may be extended in several ways: three of them are worth discussing. Barigozzi *et al.* (2019) study a two-step Generalized Dynamic Factor Model for volatilities, which also accounts for the factor structure in returns: it would be worth exploring whether our method selection criterion delivers more accurate volatility predictions within that framework. More generally, it would be interesting to study how well large dimensional factor models predict the conditional distribution of equity returns following an approach similar to that in Massacci (2015). Finally, we focused on stock returns: the analysis of bond returns predictability is high in our research agenda (see Gargano *et al.*, 2017, and references therein).

References

[1] Andersen, T.G., Bollerslev, T., Diebold, F.X., C.Vega (2007), "Real-time price discovery in global stock, bond and foreign exchange markets", *Journal of International Economics*, 73, 251–277.

[2] Aruoba, S. B., F. X. Diebold, and C. Scotti (2009), "Real-Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 27 (4), 417 – 427.

[3] Baetje, F., and L. Menkhoff (2016), "Equity Premium Prediction: Are Economic and Technical Indicators Unstable?", *International Journal of Forecasting*, 32(4), 1193 – 1207.

[4] Bai, J., and S. Ng (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70 (1), 191 – 221.

[5] Baltagi, B. H., C. Kao, and F. Wang (2017), "Identification and Estimation of a Large Factor Model with Structural Instability", *Journal of Econometrics*, 197 (1), 87 − 100.

[6] Banbura, M. D., D. Giannone, and L. Reichlin (2011), "Nowcasting," in M. P. Clements and D. F. Hendry (eds.), *The Handbook of Economic Forecasting*, 193 − 224, Oxford: Oxford University Press.

[7] Barigozzi, M., M. Hallin, and S. Soccorsi (2019), "Identification of Global and Local Shocks in International Financial Markets via General Dynamic Factor Models", *Journal of Financial Econometrics*, 17 (3), 462–494.

[8] Bekaert, G., E. Engstrom, and Y. Xing (2009), "Risk, Uncertainty and Asset Prices," *Journal of Financial Economics*, 91 (1), 59 − 82.

[9] Boivin, J., and S. Ng (2006), "Are More Data Always Better for Factor Analysis?," *Journal of Econometrics*, 132 (1), 169 − 194.

[10] Bossaerts, P., and P. Hillion (1999), "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?," *Review of Financial Studies*, 12 (2), 405 − 428.

[11] Çakmaklı, C., and D. van Dijk, (2016), "Getting the Most out of Macroeconomic Information for Predicting Excess Stock Returns", *International Journal of Forecasting*, 32(3), 650 − 668.

[12] Campbell, J. Y., and J. H. Cochrane (1999), "By Force or Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107 (2), 205 − 251.

[13] Campbell, J. Y., and S. B. Thompson (2008), "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?," *Review of Financial Studies*, 21 (4), 1509 − 1531.

[14] Chamberlain, G., and M. Rothschild (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51 (5), 1281 − 1304.

[15] Clark, T. E., and K. D. West (2007), "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics*, 138 (1), 291 − 311.

[16] Clark, T. E., and M. W. McCracken (2009), "Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts," *International Economic Review*, 50 (2), 363 − 395.

[17] Connor, G., and R. A. Korajczyk (1986), "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15 (3), 373 – 394.

[18] Connor, G., and R. A. Korajczyk (1988), "Risk and Return in an Equilibrium APT: Application of a New Test Methodology," *Journal of Financial Economics*, 21 (2), 255 – 289.

[19] Diebold, F. X. (2015), "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests," *Journal of Business & Economic Statistics*, 33 (1), 1 – 24.

[20] Elliott, G., and A. Timmermann (2016), "Forecasting in Economics and Finance," *Annual Review of Economics*, 8 (1), 81 – 110.

[21] Farmer, L., Schmidt, L., and A. Timmermann (2019), "Pockets of Predictability", SSRN working paper 3152386, `dx.doi.org/10.2139/ssrn.3152386`.

[22] Ferreira, M. A., and P. Santa-Clara (2011), "Forecasting Stock Market Returns: The Sum of the Parts Is More than the Whole," *Journal of Financial Economics*, 100 (3), 514 – 537.

[23] Fleming, J., C. Kirby, and B. Ostdiek (2001), "The Economic Value of Volatility Timing", *Journal of Finance*, 56 (1), 329 – 353.

[24] Forni, M., A. Giovannelli, M. Lippi, and S. Soccorsi (2018), "Dynamic Factor Model with Infinite-Dimensional Factor Space: Forecasting," *Journal of Applied Econometrics*, 33 (5), 625 – 642.

[25] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), "The Generalized Dynamic Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82 (4), 540 – 554.

[26] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005), "The Generalized Dynamic Factor Model: One Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100 (471), 830 – 840.

[27] Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2015), "Dynamic Factor Models with Infinite-Dimensional Factor Spaces: One-Sided Representations," *Journal of Econometrics*, 185 (2), 359 – 371.

[28] Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2017), "Dynamic Factor Models with Infinite-Dimensional Factor Space: Asymptotic analysis," *Journal of Econometrics*, 199(1), 74 − 92.

[29] Gargano, A., D. Pettenuzzo, and A. Timmermann (2017), "Bond Return Predictability: Economic Value and Links to the Macroeconomy," *Management Science*, https://doi.org/10.1287/mnsc.2017.2829.

[30] Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in D. J. Aigner and A. S. Goldberger (eds.), Latent Variables in Socio-Economic Models, North-Holland.

[31] Ghysels, E., Horan, C., Moench, E. (2018), "Forecasting through the Rearview Mirror: Data Revisions and Bond Return Predictability", *Review of Financial Studies*, 31, 678–714.

[32] Giacomini, R., and Rossi, B. (2010). "Forecast comparisons in unstable environments." *Journal of Applied Econometrics*, 25(4), 595 − 620.

[33] Giacomini, R., and H. White (2006), "Tests of Conditional Predictive Ability," *Econometrica*, 74 (6), 1545 − 1578.

[34] Giannone, D., M. Lenza, and G. E. Primiceri (2017), "Economic Predictions with Big Data: The Illusion of Sparsity," CEPR Discussion Paper No. DP12256.

[35] Gonçalves, S., M. W. McCracken, and B. Perron (2017), "Tests of Equal Accuracy for Nested Models with Estimated Factors", *Journal of Econometrics*, 198(2), 231 − 252.

[36] Goyal, A., and I. Welch (2003), "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49 (5), 639 − 654.

[37] Granger, C. W. J., and M. H. Pesaran (2000), "Economic and Statistical Measures of Forecast Accuracy," *Journal of Forecasting*, 19 (7), 426 − 459.

[38] Hallin, M., and M. Lippi (2013). "Factor Models in High-Dimensional Time Series − A Time-Domain Approach", *Stochastic Processes and their Applications*, 123(7), 2678 − 2695.

[39] Hallin, M., and R. Liška (2007), "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102 (48), 603 − 617.

[40] Hansen, P. R., and A. Timmermann (2012), "Choice of Sample Split in Out-of-Sample Forecast Evaluation," Working Paper, European University Institute and University of California, San Diego.

[41] Henkel, S. J., J. S. Martin, and F. Nardari (2011), "Time-Varying Short-Horizon Predictability," *Journal of Financial Economics*, 99 (3), 560 − 580.

[42] Inoue, A., L. Jin, and B. Rossi (2017), "Rolling Window Selection for Out-of-Sample Forecasting with Time-Varying Parameters," *Journal of Econometrics*, 196 (1), 55 − 67.

[43] Leicht, G., and J. E. Tanner (1991), "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures", *American Economic Review*, 81 (3), 580 − 590.

[44] Ludvigson, S. C., and S. Ng (2009), "The Empirical Risk-Return Relation: A Factor Analysis Approach", *Journal of Financial Economics* , 83(1), 171 − 222.

[45] Massacci, D. (2015), "Predicting the Distribution of Stock Returns: Model Formulation, Statistical Evaluation, VaR Analysis and Economic Significance," *Journal of Forecasting*, 34(3), 191 − 208.

[46] Menzly, L., T. Santos, and P. Veronesi (2004), "Understanding Predictability," *Journal of Political Economy*, 112 (1), 1 − 47.

[47] McCracken, M. W., and S. Ng (2016), "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business & Economics Statistics*, 34(4), 574 – 589.

[48] Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou (2014), "Forecasting the Equity Risk Premium: The Role of Technical Indicators," *Management Science*, 60 (7), 1772–1791.

[49] Ohno, S., and T. Ando (2018), "Stock Return Predictability: A Factor-Augmented Predictive Regression System with Shrinkage Method," *Econometric Reviews*, 37(1), 29 − 60.

[50] Pan, Z., D. Pettenuzzo, and Y. Wang (2020) "Forecasting Stock Returns: A Predictor-Constrained Approach," *Journal of Empirical Finance*, 55, 200 − 217.

[51] Paye, B. S., and A. Timmermann (2006), "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13 (3), 274 − 315.

[52] Pesaran, M. H., and A. Timmermann (1995), "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance*, 50 (4), 1201 − 1228.

[53] Pesaran, M. H., and A. Timmermann (2005), "Real-Time Econometrics," *Econometric Theory*, 21 (1), 212 – 231.

[54] Pesaran , M. H., and A. Timmermann (2007), "Selection of Estimation Window in the Presence of Breaks", *Journal of Econometrics*, 137(1) , 134–61.

[55] Pettenuzzo, D., A. Timmermann, and R. Valkanov (2014), "Forecasting Stock Returns under Economic Constraints," *Journal of Financial Economics*, 114 (3), 517 – 553.

[56] Rapach, D. E., J. K. Strauss, and G. Zhou (2010), "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy," *Review of Financial Studies*, 23 (2), 821 – 862.

[57] Rapach, D. E., and M. E. Wohar (2006), "Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns," *Journal of Financial Econometrics*, 4 (2), 238 – 274.

[58] Rapach, D. E., and G. Zhou (2013), "Forecasting Stock Returns," in G. Elliott and A. Timmermann (eds.), Handbook of Economic Forecasting, Volume 2, Part A, 329 – 383, Amsterdam: North-Holland.

[59] Sargent T., and C. A. Sims (1977), "Business Cycle Modelling without Pretending to Have too Much a Priori Economic Theory," in C. A. Sims (eds.), New Methods in Business Research, Federal Reserve Bank of Minneapolis.

[60] Soccorsi, S. (2016), "Measuring Nonfundamentalness for Structural VARs", *Journal of Economic Dynamics and Control*, 71, 86 – 101.

[61] Stock, J. H., and M. W. Watson (2002*a*), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97 (460), 1167 – 1179.

[62] Stock, J. H., and M. W. Watson (2002*b*). "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business & Economic Statistics* , 20(2), 147 – 162.

[63] Stock, J. H., and M. W. Watson (2012), "Generalized Shrinkage Methods for Forecasting Using Many Predictors," *Journal of Business & Economic Statistics*, 30 (4), 481 – 493.

[64] Timmermann, A. (2008), "Elusive Return Predictability," *International Journal of Forecasting*, 24 (1), 1 – 18.

[65] Welch, I., and A. Goyal (2008), "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21 (4), 1455 – 1508.

Table 1: Summary Statistics, 1960 - 2019

Variables	Mean	Std. Dev.	Skewness	Kurtosis
Excess Returns	0.002	0.042	-0.706	5.482
log(DP)	-3.606	0.393	-0.130	2.241
log(DY)	-3.600	0.392	-0.136	2.269
log(EP)	-2.853	0.422	-0.609	6.051
log(DE)	-0.753	0.305	2.792	18.807
SVAR	0.002	0.004	10.455	145.403
BM	0.490	0.257	0.812	2.802
NTIS	0.010	0.020	-0.644	3.127
TBL	0.045	0.032	0.736	3.890
LTY	0.063	0.027	0.713	3.143
LTR	0.006	0.029	0.436	5.661
TMS	0.018	0.014	-0.259	2.797
DFY	0.010	0.004	1.797	7.586
DFR	0.000	0.015	-0.393	9.409
INFL	0.003	0.004	0.050	6.071

Notes. This table reports summary statistics for excess returns on the S&P 500 index in excess of the treasury bill rate and for the following 14 predictors proposed in Welch and Goyal (2008): log dividend-price ratio (log (DP)), log dividend-yield (log (DY)), log earnings-price ratio (log (EP)), log dividend-payout ratio (log (DE)), stock variance (SVAR), book-to-market ratio (BM), net equity expansion (NTIS), treasury bill rate (TBL), long-term yield (LTY), long-term return (LTR), term spread (TMS), default yield spread (DFY), default return spread (DFR), inflation (INFL). The sample period is $1960 - 2019$.

Table 2: Out-Of-Sample Forecast Performance, Recursive Window, 1970 - 2019

Panel A: Welch and Goyal (2008) Data

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r = 1$	-1.21	-0.58	-2.73	$q = 1$ $r = 1$	-0.82	-0.35	-1.94	$q = 1$	0.98***	0.81**	1.39**
$r = 2$	-0.16*	0.01**	-0.56	$r = 2$	-1.79	-0.06	-5.98				
$r = 3$	-2.73	-1.67*	-5.29	$r = 3$	-2.36	0.28**	-8.74				
				$q = 2$ $r = 2$	-0.07	0.14	-0.57	$q = 2$	0.17	0.5*	-0.62
				$r = 3$	-0.88	-0.42*	-1.98				

Panel B: FRED-MD Data

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r = 1$	-0.73	0.23*	-3.06	$q = 1$ $r = 1$	-0.71	0.02	-2.47	$q = 1$	0.58*	0.39*	1.03
$r = 2$	-0.04**	0.02***	-0.18	$r = 2$	1.30***	0.71**	2.74*				
				$r = 3$	1.30***	0.63**	2.91*				
$r = 5$	-1.48**	-1.91*	-0.45	$r = 5$	0.46**	0.19**	1.12				
$r = 15$	-4.44*	-4.57*	-4.13	$r = 15$	-0.79*	-1.73	1.48*				
				$q = 3$ $r = 5$	-0.2**	-0.88**	1.43*	$q = 3$	0.46*	0.33*	0.79
				$r = 15$	-2.33*	-3.04	-0.61				
				$q = 5$ $r = 5$	-0.25**	-0.97**	1.48*	$q = 5$	0.87**	0.58**	1.55
				$r = 15$	-2.95	-3.61	-1.34				

Panel C: FRED-MD Data, Lasso 25

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r = 1$	-0.88	-0.38*	-2.11	$q = 1$ $r = 1$	-0.59	-0.12*	-1.72	$q = 1$	-0.20	0.07	-0.85
$r = 5$	-2.87*	-3.25	-1.95	$r = 5$	-1.19**	-1.19**	-1.19				
$r = 15$	-7.27	-7.95	-5.64	$r = 15$	-1.84*	-2.79	0.46*				
				$q = 3$ $r = 5$	-2.44	-2.01*	-3.49	$q = 3$	-1.83	-2.61	0.04
				$r = 15$	-6.09	-6.61	-4.82				
				$q = 5$ $r = 5$	-2.34	-2.39	-2.21	$q = 5$	-3.08	-4.49	0.34
				$r = 15$	-5.75	-6.18	-4.72				

Panel D: FRED-MD Data, Lasso 50

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r = 1$	-0.71	0.14*	-2.75	$q = 1$ $r = 1$	-0.16	0.7**	-2.21	$q = 1$	-0.86	-0.19	-2.48
$r = 5$	-2.56*	-3.00	-1.49	$r = 5$	-0.9*	-0.82*	-1.08				
$r = 15$	-6.44	-6.43	-6.45	$r = 15$	-1.06*	-1.38	-0.27				
				$q = 3$ $r = 5$	-2.17	-1.72*	-3.25	$q = 3$	-0.80	-0.55	-1.41
				$r = 15$	-3.72	-5.02	-0.59				
				$q = 5$ $r = 5$	-1.76*	-1.63*	-2.06	$q = 5$	-0.62	-1.77	2.14*
				$r = 15$	-2.93*	-5.00	2.07*				

Panel E: FRED-MD Data, Lasso 75

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r = 1$	-0.63	0.66**	-3.74	$q = 1$ $r = 1$	-0.49	0.79**	-3.57	$q = 1$	1.05**	0.65**	2.00
$r = 5$	-1.87*	-1.42**	-2.95	$r = 5$	0.02**	0.26**	-0.54				
$r = 15$	-5.70	-5.31	-6.66	$r = 15$	-1.02*	-0.69*	-1.83				
				$q = 3$ $r = 5$	-0.64**	-0.54**	-0.88	$q = 3$	-0.06	0.32*	-0.97
				$r = 15$	-2.58*	-4.01	0.86				
				$q = 5$ $r = 5$	-0.98**	-0.64**	-1.79	$q = 5$	-0.37	-0.18	-0.81
				$r = 15$	-3.44	-4.96	0.24				

Notes. This table shows the out-of-sample R^2 (%) of prediction models for the monthly excess return $\rho_{\tau+1}$ using the recursive window estimation scheme. The factor models are Stock and Watson (2002) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ), described in Section 2. Factors are estimated from the Welch and Goyal (2008) dataset in Panel A. Factors are extracted from the FRED-MD dataset detailed in MacCracken and Ng (2015) in Panel B, and from 25, 50 and 75 series selected with a LASSO estimator in Panels C, D and E, respectively. Statistical significance is assessed through the Clarke and West (2007) CW statistic: *, ** and *** denote significance at 10%, 5% and 1%, respectively. The sample period is 1970 − 2019.

Table 3: Resume of small to medium LASSO Results, FRED-MD Data, 1970 - 2019

		LASSO, 25	LASSO, 50	LASSO, 75	full FRED-MD dataset
Panel A: Recursive Window	SW	-0.88	-0.71	-0.63	-0.04^{**}
	FHLR	-0.59	-0.16	0.08^{**}	1.30^{***}
	FHLZ	-0.20	-0.22	1.05^{**}	0.87^{**}
Panel B: Rolling Window	SW	-2.25	-2.32	-0.30^{*}	-1.13
	FHLR	-1.33	-1.03	-0.43^{**}	0.67^{***}
	FHLZ	-0.63	1.57^{**}	1.04^{**}	1.96^{***}

Notes. This table shows the out-of-sample R^2 (%) of prediction models for the monthly excess return $\rho_{\tau+1}$ using recursive window (Panel A) and rolling window (Panel B) estimation schemes. The factor models are Stock and Watson (2002) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ), described in Section 2. For each model, we report the most accurate specification among all results obtained for $r = 1, 2, 3, 4, 5, 10, 15$, and $q = 1, 2, 3, 4, 5$. Factors are extracted from 25, 50 and 75 series selected at each point in time with a LASSO estimator applied to the FRED-MD dataset detailed in MacCracken and Ng (2015). Statistical significance is assessed through the Clarke and West (2007) CW statistic: *, ** and *** denote significance at 10%, 5% and 1%, respectively. The sample period is 1970 − 2019.

Table 4: Out-Of-Sample Results, FRED-MD Data, MSPE, 1970 - 2019

Panel A: Recursive Window

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r = 1$	0.1938	0.1575	0.4196	$q = 1$ $r = 1$	0.1937	0.1579	0.4172	$q = 1$	0.1913	0.1573	0.4030
$r = 5$	0.1952	0.1609	0.4089	$r = 5$	0.1915	0.1576	0.4026				
$r = 15$	0.2009	0.1651	0.4239	$r = 15$	0.1939	0.1606	0.4011				
				$q = 3$ $r = 5$	0.1928	0.1593	0.4013	$q = 3$	0.1915	0.1574	0.4039
				$r = 15$	0.1968	0.1627	0.4096				
				$q = 5$ $r = 5$	0.1929	0.1594	0.4011	$q = 5$	0.1907	0.1570	0.4008
				$r = 15$	0.1980	0.1636	0.4126				

Panel B: Rolling Window

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r = 1$	0.1952	0.1591	0.4199	$q = 1$ $r = 1$	0.1964	0.1594	0.4267	$q = 1$	0.1913	0.1591	0.3920
$r = 5$	0.2013	0.1692	0.4010	$r = 5$	0.1938	0.1639	0.3798				
$r = 15$	0.2323	0.1932	0.4756	$r = 15$	0.2012	0.1703	0.3941				
				$q = 3$ $r = 5$	0.1951	0.1648	0.3842	$q = 3$	0.1907	0.1583	0.3923
				$r = 15$	0.2045	0.1713	0.4116				
				$q = 5$ $r = 5$	0.1954	0.1650	0.3851	$q = 5$	0.1892	0.1578	0.3848
				$r = 15$	0.2055	0.1735	0.4053				
BN	0.2029	0.1696	0.4105	HL, BN	0.1946	0.1641	0.3844	HL	0.1901	0.1583	0.3884

Notes. This table shows the mean squared prediction error (MSPE $\times 100$) of prediction models for the monthly excess return $\rho_{\tau+1}$ using recursive window (Panel A) and rolling window (Panel B) estimation schemes. The factor models are Stock and Watson (2002) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ), described in Section 2. Factors are extracted from the FRED-MD dataset detailed in MacCracken and Ng (2015). BN and HL denote Bai and Ng (2002), and Hallin and Liška (2007) model selection criteria, respectively. The sample period is $1970 - 2019$.

Table 5: Out-Of-Sample Forecast Performance, Method Selection Criterion, 1970 - 2019

PANEL A: MSPE

Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
0.1924	0.1952	0.1931	0.1911	0.1930	0.2029	0.1946	0.1901	0.1928	0.1986	0.1911	0.1897

PANEL B: Out-of-Sample R^2

Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-1.48**	-0.36**	0.68**	-0.33	-5.47**	-1.14**	1.19**	-0.24	-2.87**	0.65***	1.41**

Panel C: Portfolio Choice, Δ %(ann.) (%), $\gamma = 3$

Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	1.07	-0.29	-1.54	0.99	0.35	1.27	1.38	-0.03	0.56	0.80	0.013

Panel D: Portfolio Choice, Δ %(ann.) (%), $\gamma = 4$

Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-0.30	-1.00	-1.60	0.58	-0.16	0.62	1.00	-0.02	0.19	0.48	0.17

Panel E: Portfolio Choice, Δ %(ann.) (%), $\gamma = 5$

Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-1.33	-1.54	-1.39	0.16	-0.73	-0.13	0.60	-0.02	0.08	0.23	0.21

Panel F: Portfolio Choice, Δ %(ann.) (%), $\gamma = 10$

Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-2.85	-2.62	-0.68	-0.20	-2.94	-1.75	-0.04	-0.01	-0.23	0.04	0.12

Notes. This table shows mean squared prediction error (MSPE ×100), out-of-sample R^2 (%) and annualized utility gain (%) of prediction models for the monthly excess return $\rho_{\tau+1}$. The prevailing mean PM is defined in (13). Stock and Watson (2002*a*) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ) factor models are described in Section 2. Factors are extracted from the FRED-MD dataset detailed in MacCracken and Ng (2015). Under recursive window, $r = 5$ and $q = 4$ static and dynamic factors, respectively, are selected as detailed in Section 5.2.1. Under rolling window, the number of static and dynamic factors is selected according to Bai and Ng (2002), and Hallin and Liška (2007) model selection criteria, respectively. The method selection criterion allows to switch between estimation windows using the procedure described in Section 5.3.2. Statistical significance is assessed through the Clarke and West (2007) CW statistic: *, ** and *** denote significance at 10%, 5% and 1%, respectively. The sample period is $1970 - 2019$.

Table 6: Real-time vs pseudo real-time results

PSEUDO REAL TIME

PANEL A: MSPE based on FRED Dataset

Recursive Window (1992-2019)				Rolling Window (1992-2019)				Method Selection (2009-2019)			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
0.1664	0.1692	0.1676	0.1655	0.1676	0.1819	0.1745	0.1668	0.1548	0.1769	0.1663	0.1631

PANEL B: Out-of-Sample R^2 based on FRED Dataset

Recursive Window (1992-2019)				Rolling Window (1992-2019)				Method Selection (2009-2019)			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-1.65	-0.77	0.53	-0.74	-9.3	-4.89	-0.77	-3.07	-17.76	-10.81	-8.71

REAL TIME

PANEL C: MSPE based on ALFRED Dataset

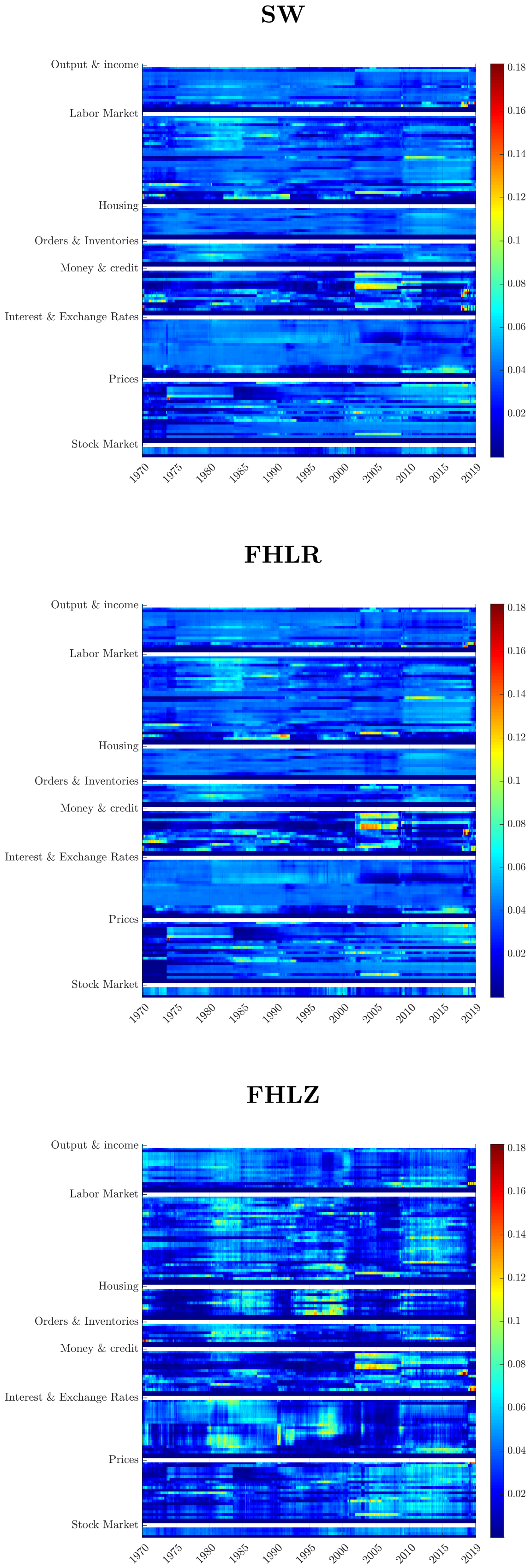
Recursive Window (1992-2019)				Rolling Window (1992-2019)				Method Selection (2009-2019)			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
0.1664	0.1693	0.1685	0.1667	0.1676	0.1682	0.1688	0.1629	0.1593	0.1650	0.1627	0.1591

PANEL D: Out-of-Sample R^2 based on ALFRED Dataset

Recursive Window (1992-2019)				Rolling Window (1992-2019)				Method Selection (2009-2019)			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-1.77	-1.31	-0.22	-0.74	-1.10	-1.45	2.01**	-1.83	-5.52	-4.04	-1.71

Notes. This table shows mean squared prediction error (MSPE $\times 100$) and out-of-sample R^2 (%) of prediction models for the monthly excess return $\rho_{\tau+1}$. The prevailing mean PM is defined in (13). Stock and Watson (2002*a*) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ) factor models are described in Section 2. Factors are extracted from the FRED-MD dataset detailed in MacCracken and Ng (2015). Under recursive window, $r = 5$ and $q = 4$ static and dynamic factors, respectively, are selected as detailed in Section 5.2.1. Under rolling window, the number of static and dynamic factors is selected according to Bai and Ng (2002), and Hallin and Liška (2007) model selection criteria, respectively. The method selection criterion allows to switch between estimation windows using the procedure described in Section 5.3.2. Statistical significance is assessed through the Clarke and West (2007) CW statistic: *, ** and *** denote significance at 10%, 5% and 1%, respectively. The sample period is $1992 - 2019$.

Figure 1: Degree of commonality

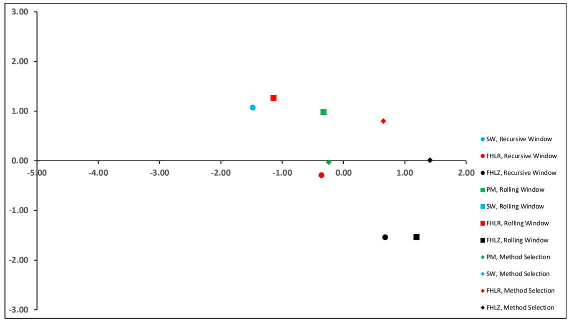


Notes. The heatmaps above report the (rolling) degree of commonality estimated by the three factor models as in equation (12)

Figure 2: Statistical and Economic Out-of-Sample Forecast Accuracy

Panel A: $\gamma = 3$

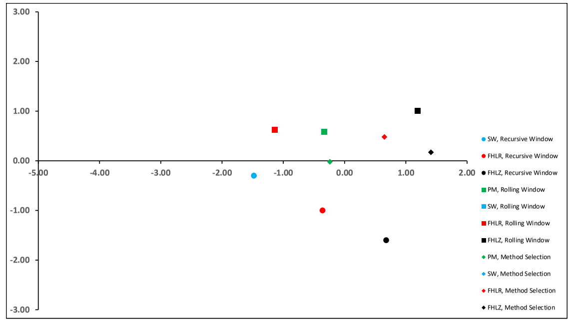
Δ (ann.) (%)



out-of-sample R^2 (%)

Panel B: $\gamma = 4$

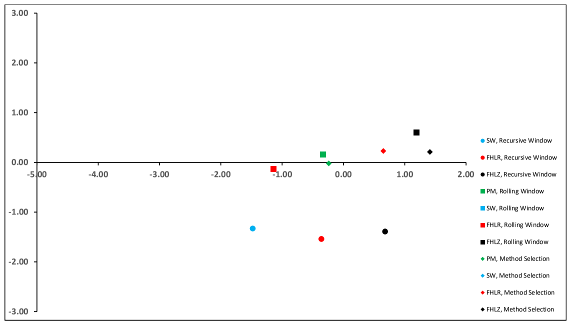
Δ (ann.) (%)



out-of-sample R^2 (%)

Panel C: $\gamma = 5$

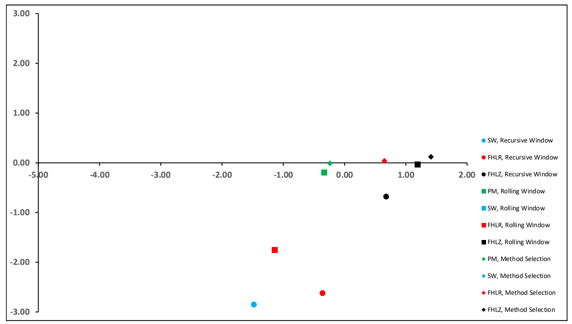
Δ (ann.) (%)



out-of-sample R^2 (%)

Panel D: $\gamma = 10$

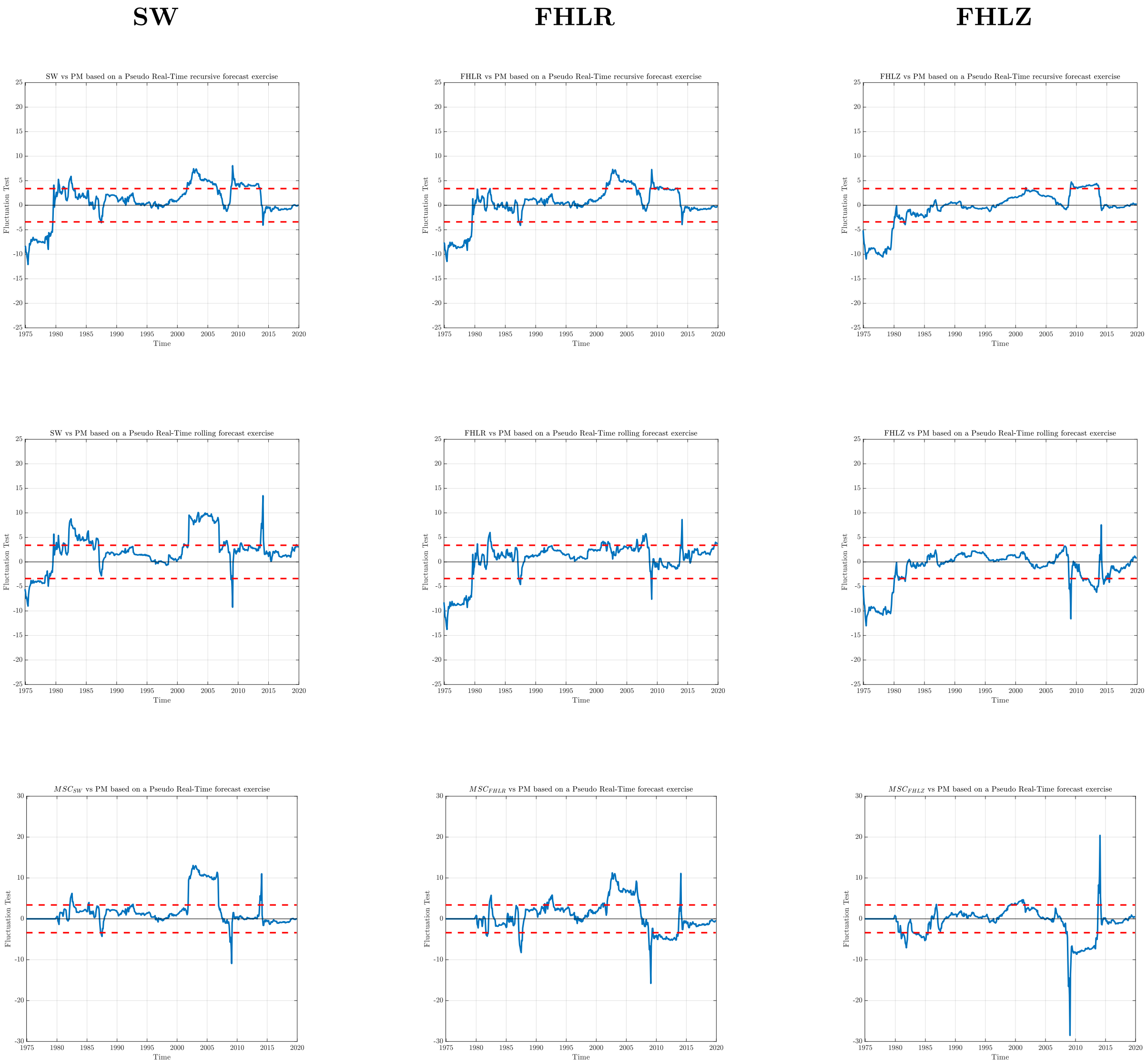
Δ (ann.) (%)



out-of-sample R^2 (%)

Notes. This figure illustrates the link between statistical and economic out-of-sample forecast accuracy as measured by out-of-sample R^2 (%) and annualized utility gain (%), respectively, of prediction models for the monthly excess return $\rho_{\tau+1}$. γ denotes the relative risk aversion parameter. Stock and Watson (2002) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ) factor models are described in Section 2. Recursive (Rec) and rolling (Rol) window estimation schemes are described in Section 5.1.2. The method selection criterion (Sel) allows to switch between estimation windows using the procedure described in Section 5.3.2. The sample period is 1970 – 2019.

Figure 3: GR test: SW, FHLR, FHLZ vs PM benchmark.



Notes. Forecasting methods are: recursive (top), rolling (centre) and method selection (*MSC*, bottom). GR test statistics (blue) within 5% confidence bands (red dashed); the smoothing adopted is 60 data points.

Online Appendix

Rolling Window Results

Table A1: Out-Of-Sample Forecast Performance, Rolling Window, 1970 - 2019

Panel A: Welch and Goyal (2008) Data

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r=1$	-0.90	0.01	-3.10	$q=1$ $r=1$	-3.32	-1.57	-7.60	$q=1$	1.07**	1.09**	1.03
$r=2$	-0.44**	-0.24**	-0.93	$r=2$	-2.35	-1.83	-3.62				
$r=3$	-3.19	-2.18*	-5.67	$r=3$	-4.54	-2.73	-8.94				
				$q=2$ $r=2$	-3.55	-3.71	-3.15	$q=2$	0.40*	0.97**	-0.98
				$r=3$	-4.19	-3.03	-7.01				

Panel B: FRED-MD Data

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r=1$	-1.13	-0.15*	-3.51	$q=1$ $r=1$	-1.76	-0.36	-5.18	$q=1$	0.88**	-0.15	3.37**
$r=5$	-4.30**	-6.53	1.14**	$r=5$	-0.41**	-3.19	6.37**				
$r=15$	-20.35	-21.62	-17.24	$r=15$	-4.27	-7.18	2.84*				
				$q=2$ $r=2$	0.67***	-0.15**	2.66*				
				$q=3$ $r=5$	-1.09**	-3.71	5.29**	$q=3$	1.22**	0.37**	3.29*
				$r=15$	-5.97	-7.81	-1.48				
				$q=5$ $r=5$	-1.27**	-3.86	5.06**	$q=5$	1.96***	0.66**	5.14**
				$r=15$	-6.50	-9.20	0.09				
BN	-5.12**	-6.73	-1.19*	HL, BN	-0.81**	-3.29	5.24**	HL	1.51**	0.38**	4.26*

Panel C: FRED-MD Data, Lasso 25

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r=1$	-2.25	-2.38	-1.93	$q=1$ $r=1$	-1.33	-1.82	-0.11	$q=1$	-0.96	-1.35	-0.01
$r=5$	-8.12	-7.08	-10.65	$r=5$	-2.79	-5.35	3.43*				
$r=15$	-25.94	-28.63	-19.39	$r=15$	-2.88*	-7.58	8.58***				
				$q=3$ $r=5$	-2.79*	-4.55	1.51*	$q=3$	-0.76	-2.67	3.92*
				$r=15$	-8.68	-13.82	3.87**				
				$q=5$ $r=5$	-1.62**	-3.13	2.06*	$q=5$	-1.14	-2.70	2.68*
				$r=15$	-10.24	-14.38	-0.16				

Panel D: FRED-MD Data, Lasso 50

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r=1$	-2.32	-1.12	-5.23	$q=1$ $r=1$	-1.03	-0.09	-3.33	$q=1$	0.39	-0.61	2.83**
$r=5$	-7.88	-8.03	-7.53	$r=5$	-3.24	-3.89	-1.66				
$r=15$	-18.66	-19.75	-16.00	$r=15$	-4.12	-6.87	2.58*				
				$q=3$ $r=5$	-1.89**	-2.89	0.56*	$q=3$	1.12**	-0.34	4.7**
				$r=15$	-5.58*	-8.62	1.85*				
				$q=5$ $r=5$	-1.99**	-3.18	0.91*	$q=5$	0.72**	-0.95	4.79**
				$r=15$	-5.24**	-8.62	2.99*				

Panel E: FRED-MD Data, Lasso 75

SW				FHLR				FHLZ			
Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions	Model	Full Sample	Expansion	Recessions
$r=1$	-0.3*	-0.26*	-0.42	$q=1$ $r=1$	-0.52	-0.06	-1.63	$q=1$	0.71**	-0.14	2.77**
$r=5$	-5.58*	-7.10	-1.87	$r=5$	-2.61	-2.99	-1.68				
$r=15$	-17.02	-18.95	-12.32	$r=15$	-2.21	-4.11	2.43*				
				$q=3$ $r=5$	-0.43**	-3.33	6.64**	$q=3$	0.89**	-0.06	3.21*
				$r=15$	-1.93**	-4.56*	4.47*				
				$q=5$ $r=5$	-0.28**	-3.32	7.13**	$q=5$	1.04**	-0.20	4.06*
				$r=15$	-5.02	-8.76	4.1*				

Notes. This table shows the out-of-sample R^2 (%) of prediction models for the monthly excess return $\rho_{\tau+1}$ using the rolling window estimation scheme. The factor models are Stock and Watson (2002) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ), described in Section 2. Factors are estimated from the Welch and Goyal (2008) (WG) dataset in Panel A. Factors are extracted from the FRED-MD dataset detailed in MacCracken and Ng (2015) in Panel B, and from 25, 50 and 75 series selected with a LASSO estimator in Panels C, D and E, respectively. Statistical significance is assessed through the Clarke and West (2007) CW statistic: *, ** and *** denote significance at 10%, 5% and 1%, respectively. BN and HL denote Bai and Ng (2002), and Hallin and Liška (2007) model selection criteria, respectively. The sample period is $1970-2019$.

FRED-MD Data

We adopt the balanced version of FRED-MD dataset discarding the series with missing values at beginning of the sample. These are: PERMIT, PERMITNE, PERMITMW, PERMITS, PERMITW, ACOGNO, ANDENOx, TWEXMMTH, UMCSENTx.

Letting x_t be a raw series, the transformations adopted are:

(1) no transformation;

(2) Δx_t;

(3) $\Delta^2 x_t$;

(4) $\Delta log(x_t)$;

(5) $log(x_t)$;

(6) $\Delta^2 log(x_t)$;

(7) $\Delta \left(\frac{x_t}{x_{t-1}-1}\right) log(x_t)$;

Table A2: List of FRED-MD time series

	mnemonic	description	tcode
1	RPI	Real Personal Income	5
2	W875RX1	RPI ex. Transfers	5
3	INDPRO	IP Index	5
4	IPFPNSS	IP: Final Products and Supplies	5
5	IPFINAL	IP: Final Products	5
6	IPCONGD	IP: Consumer Goods	5
7	IPDCONGD	IP: Durable Consumer Goods	5
8	IPNCONGD	IP: Nondurable Consumer Goods	5
9	IPBUSEQ	IP: Business Equipment	5
10	IPMAT	IP: Materials	5
11	IPDMAT	IP: Durable Materials	5
12	IPNMAT	IP: Nondurable Materials	5
13	IPMANSICS	IP: Manufacturing	5
14	IPB51222S	IP: Residential Utilities	5
15	IPFUELS	IP: Fuels	5
16	NAPMPI	ISM Manufacturing: Production	1
17	CAPUTLB00004S	Capacity Utilization: Manufacturing	2
1	HWI	Help-Wanted Index for US	2
2	HWIURATIO	Help Wanted to Unemployed ratio	2
3	CLF16OV	Civilian Labor Force	5
4	CE16OV	Civilian Employment	5
5	UNRATE	Civilian Unemployment Rate	2
6	UEMPMEAN	Average Duration of Unemployment	2
7	UEMPLT5	Civilians Unemployed <5 Weeks	5
8	UEMP5TO14	Civilians Unemployed 5-14 Weeks	5
9	UEMP15OV	Civilians Unemployed >15 Weeks	5
10	UEMP15T26	Civilians Unemployed 15-26 Weeks	5
11	UEMP27OV	Civilians Unemployed >27 Weeks	5
12	CLAIMSx	Initial Claims	5
13	PAYEMS	All Employees: Total nonfarm	5
14	USGOOD	All Employees: Goods-Producing	5
15	CES1021000001	All Employees: Mining and Logging	5
16	USCONS	All Employees: Construction	5
17	MANEMP	All Employees: Manufacturing	5
18	DMANEMP	All Employees: Durable goods	5
19	NDMANEMP	All Employees: Nondurable goods	5
20	SRVPRD	All Employees: Service Industries	5
21	USTPU	All Employees: TT&U	5
22	USWTRADE	All Employees: Wholesale Trade	5
23	USTRADE	All Employees: Retail Trade	5
24	USFIRE	All Employees: Financial Activities	5
25	USGOVT	All Employees: Government	5
26	CES0600000007	Hours: Goods-Producing	1
27	AWOTMAN	Overtime Hours: Manufacturing	2
28	AWHMAN	Hours: Manufacturing	1
29	NAPMEI	ISM Manufacturing: Employment	1
30	CES0600000008	Ave. Hourly Earnings: Goods	6
31	CES2000000008	Ave. Hourly Earnings: Construction	6
32	CES3000000008	Ave. Hourly Earnings: Manufacturing	6
1	HOUST	Starts: Total	4
2	HOUSTNE	Starts: Northeast	4
3	HOUSTMW	Starts: Midwest	4
4	HOUSTS	Starts: South	4
5	HOUSTW	Starts: West	4

- Continued on next page -

Table A2 – continued from previous page

1	DPCERA3M086SBEA	Real PCE	5
2	CMRMTSPLx	Real M&T Sales	5
3	RETAILx	Retail and Food Services Sales	5
4	NAPM	ISM: PMI Composite Index	1
5	NAPMNOI	ISM: New Orders Index	1
6	NAPMSDI	ISM: Supplier Deliveries Index	1
7	NAPMII	ISM: Inventories Index	1
8	AMDMNOx	Orders: Durable Goods	5
9	AMDMUOx	Unfilled Orders: Durable Goods	5
10	BUSINVx	Total Business Inventories	5
11	ISRATIOx	Inventories to Sales Ratio	2
1	M1SL	M1 Money Stock	6
2	M2SL	M2 Money Stock	6
3	M3SL	MABMM301USM189S in FRED, M3 for the United States	6
4	M2REAL	Real M2 Money Stock	5
5	AMBSL	St. Louis Adjusted Monetary Base	6
6	TOTRESNS	Total Reserves	6
7	NONBORRES	Nonborrowed Reserves	6
8	BUSLOANS	Commercial and Industrial Loans	6
9	REALLN	Real Estate Loans	1
10	NONREVSL	Total Nonrevolving Credit	6
11	CONSPI	Credit to PI ratio	2
12	MZMSL	MZM Money Stock	6
13	DTCOLNVHFNM	Consumer Motor Vehicle Loans	6
14	DTCTHFNM	Total Consumer Loans and Leases	6
15	INVEST	Securities in Bank Credit	6
1	FEDFUNDS	Effective Federal Funds Rate	2
2	CP3M	3-Month AA Comm. Paper Rate	2
3	TB3MS	3-Month T-bill	2
4	TB6MS	6-Month T-bill	2
5	GS1	1-Year T-bond	2
6	GS5	5-Year T-bond	2
7	GS10	10-Year T-bond	2
8	AAA	Aaa Corporate Bond Yield	2
9	BAA	Baa Corporate Bond Yield	2
10	COMPAPFF	CP - FFR spread	1
11	TB3SMFFM	3 Mo. - FFR spread	1
12	TB6SMFFM	6 Mo. - FFR spread	1
13	T1YFFM	1 yr. - FFR spread	1
14	T5YFFM	5 yr. - FFR spread	1
15	T10YFFM	10 yr. - FFR spread	1
16	AAAFFM	Aaa - FFR spread	1
17	BAAFFM	Baa - FFR spread	1
18	EXSZUS	Switzerland / U.S. FX Rate	5
19	EXJPUS	Japan / U.S. FX Rate	5
20	EXUSUK	U.S. / U.K. FX Rate	5
21	EXCAUS	Canada / U.S. FX Rate	5
1	PPIFGS	PPI: Finished Goods	6
2	PPIFCG	PPI: Finished Consumer Goods	6
3	PPIITM	PPI: Intermediate Materials	6
4	PPICRM	PPI: Crude Materials	6
5	oilprice	Crude Oil Prices: WTI	6
6	PPICMM	PPI: Commodities	6
7	NAPMPRI	ISM Manufacturing: Prices	1
8	CPIAUCSL	CPI: All Items	6
9	CPIAPPSL	CPI: Apparel	6
10	CPITRNSL	CPI: Transportation	6
11	CPIMEDSL	CPI: Medical Care	6
12	CUSR0000SAC	CPI: Commodities	6
13	CUUR0000SAD	CPI: Durables	6
14	CUSR0000SAS	CPI: Services	6
15	CPIULFSL	CPI: All Items Less Food	6
16	CUUR0000SA0L2	CPI: All items less shelter	6
17	CUSR0000SA0L5	CPI: All items less medical care	6
18	PCEPI	PCE: Chain-type Price Index	6
19	DDURRG3M086SBEA	PCE: Durable goods	6
20	DNDGRG3M086SBEA	PCE: Nondurable goods	6
21	DSERRG3M086SBEA	PCE: Services	6
1	S&P 500	S&P: Composite	5
2	S&P: indust	S&P: Industrials	5
3	S&P div yield	S&P: Dividend Yield	2
4	S&P PE ratio	S&P: Price-Earnings Ratio	5

ALFRED data

Table A3: List of ALFRED time series

Mnemonic	Variable Description	TCode	Start Date
AWHMAN	Avg Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	4	11/1/1964
AWHNONAG	Avg Weekly Hours Of Production And Nonsupervisory Employees: Total private	4	5/1/1970
AWOTMAN	Avg Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing	4	8/1/1966
CE16OV	Civilian Employment	4	12/1/1964
CLF16OV	Civilian Labor Force	4	11/1/1964
CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items	4	6/1/1972
CURRDD	Currency Component of M1 Plus Demand Deposits	4	11/1/1964
CURRSL	Currency Component of M1	4	11/1/1964
DEMDEPSL	Demand Deposits at Commercial Banks	4	9/1/1964
DMANEMP	All Employees: Durable goods	4	11/1/1964
DSPI	Disposable Personal Income	4	1/1/1980
DSPIC96	Real Disposable Personal Income	4	2/1/1980
HOUST	Housing Starts: Total: New Privately Owned Housing Units Started	4	12/1/1964
HOUST1F	Privately Owned Housing Starts: 1-Unit Structures	4	2/1/1972
HOUST2F	Housing Starts: 2-4 Units	4	2/1/1973
INDPRO	Industrial Production Index	4	11/1/1964
M1SL	M1 Money Stock	4	12/1/1979
M2SL	M2 Money Stock	4	12/1/1979
MANEMP	All Employees: Manufacturing	4	11/1/1964
NDMANEMP	All Employees: Nondurable goods	4	11/1/1964
OCDSL	Other Checkable Deposits	4	2/1/1981
PAYEMS	All Employees: Total nonfarm	4	11/1/1964
PCE	Personal Consumption Expenditures	4	12/1/1979
PCEC96	Real Personal Consumption Expenditures	4	3/1/1980
PCEDG	Personal Consumption Expenditures: Durable Goods	4	12/1/1979
PCEDGC96	Real Personal Consumption Expenditures: Durable Goods	4	3/1/1980
PCEND	Personal Consumption Expenditures: Nondurable Goods	4	12/1/1979
PCENDC96	Real Personal Consumption Expenditures: Nondurable Goods	4	3/1/1980
PCES	Personal Consumption Expenditures: Services	4	12/1/1979
PCESC96	Real Personal Consumption Expenditures: Services	4	3/1/1980
PFCGEF	Producer Price Index: Finished Consumer Goods Excluding Foods	4	1/1/1982
PI	Personal Income	4	2/1/1966
PPICFF	Producer Price Index: Crude Foodstus \& Feedstus	4	1/1/1982
PPICPE	Producer Price Index: Finished Goods: Capital Equipment	4	1/1/1978
PPICRM	Producer Price Index: Crude Materials for Further Processing	4	3/1/1978
PPIFCF	Producer Price Index: Finished Consumer Foods	4	1/1/1982
PPIFGS	Producer Price Index: Finished Goods	4	1/1/1982
PPIIFF	Producer Price Index: Intermediate Foods \& Feeds	4	1/1/1982
PPIITM	Producer Price Index: Intermediate Materials: Supplies \& Components	4	3/1/1978
SAVINGSL	Savings Deposits - Total	4	12/1/1979
SRVPRD	All Employees: Service-Providing Industries	4	9/1/1971
STDCBSL	Small Time Deposits at Commercial Banks	4	12/1/1979
STDSL	Small Time Deposits - Total	4	12/1/1979
STDTI	Small Time Deposits at Thrift Institutions	4	12/1/1979
SVGCBSL	Savings Deposits at Commercial Banks	4	12/1/1979
SVGTI	Savings Deposits at Thrift Institutions	4	12/1/1979
SVSTCBSL	Savings and Small Time Deposits at Commercial Banks	4	12/1/1979
SVSTSL	Savings and Small Time Deposits - Total	4	12/1/1979
TCDSL	Total Checkable Deposits	4	3/1/1981
UEMP15OV	Civilians Unemployed - 15 Weeks \& Over	4	11/1/1964
UEMP15T26	Civilians Unemployed for 15-26 Weeks	4	1/1/1982
UEMP27OV	Civilians Unemployed for 27 Weeks and Over	4	1/1/1966
UEMP5TO14	Civilians Unemployed for 5-14 Weeks	4	11/1/1964
UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	4	11/1/1964
UEMPMEAN	Average (Mean) Duration of Unemployment	4	1/1/1972
UEMPMED	Median Duration of Unemployment	4	1/1/1982
UNEMPLOY	Unemployed	4	12/1/1964
UNRATE	Civilian Unemployment Rate	2	2/1/1960
USCONS	All Employees: Construction	4	12/1/1964
USFIRE	All Employees: Financial Activities	4	12/1/1964
USGOOD	All Employees: Goods-Producing Industries	4	9/1/1971
USGOVT	All Employees: Government	4	12/1/1964
USMINE	All Employees: Mining and logging	4	12/1/1964
USPRIV	All Employees: Total Private Industries	4	8/1/1971
USSERV	All Employees: Other Services	4	12/1/1964
USTPU	All Employees: Trade, Transportation \& Utilities	4	12/1/1964
USTRADE	All Employees: Retail Trade	4	12/1/1964
USWTRADE	All Employees: Wholesale Trade	4	12/1/1964